

3341 GTTCAGCTGACATTTCTCAACACACCACTCACCGATGTATTCGTTGCCAGTGGAA 2400
3342 GTTCAGCTGACATTTCTCAACACACCACTCACCGATGTATTCGTTGCCAGTGGAA 2400
3401 CATACAGCTGAAATTTGGAAATAAATTTGTTTCTTGGCCAGGAATACTAGACTG 2400
3402 CATACAGCTGAAATTTGGAAATAAATTTGTTTCTTGGCCAGGAATACTAGACTG 2400
3461 ACTTTGATGCTCCCAACACATACCCAGTGTAAACACAGAGATGTGGAGGGAGCTGG 2520
3462 ACTTTGATGCTCCCAACACATACCCAGTGTAAACACAGAGATGTGGAGGGAGCTGG 2520
3521 GAGATTTCAGTGGTACATTTGAATTCCTCCAACTACCCAGGCAATTAACCCAGCAACCG 2580
3522 GAGATTTCAGTGGTACATTTGAATTCCTCCAACTACCCAGGCAATTAACCCAGCAACCG 2580
3581 AGTGTAGTGGACCATCAACCCACCCCAAGCGCGCATCTGATGCTGCTCCCTGAGA 2640
3582 AGTGTAGTGGACCATCAACCCACCCCAAGCGCGCATCTGATGCTGCTCCCTGAGA 2640
3641 TCTTCTGCTCCATAGAGGAGCTGTGGGAGCTATCTGCTGATGCGGAAACCTCTTCAT 2700
3642 TCTTCTGCTCCATAGAGGAGCTGTGGGAGCTATCTGCTGATGCGGAAACCTCTTCAT 2700
3701 CCAATTCTGTGACACATATGAACCTGCGAGACCTACGAAACCGCCCATCGCTTCACCT 2760
3702 CCAATTCTGTGACACATATGAACCTGCGAGACCTACGAAACCGCCCATCGCTTCACCT 2760
3761 CCAAGTCAAAAGAGCTGTGGATTCAAGTCCAACTCAAGGGAACAGCGCTAGAGGCT 2820
3762 CCAAGTCAAAAGAGCTGTGGATTCAAGTCCAACTCAAGGGAACAGCGCTAGAGGCT 2820
3821 TCCAGTCCCATACGTGACATATGATGAGGAGCTTACAGGAACTCATTTGAAGACATAGTC 2880
3822 TCCAGTCCCATACGTGACATATGATGAGGAGCTTACAGGAACTCATTTGAAGACATAGTC 2880
3881 GAGATGCGAGCTCTATGATCTGAAACATCAGGAAATCTTAAGATAGAACTTA 2940
3882 GAGATGCGAGCTCTATGATCTGAAACATCAGGAAATCTTAAGATAGAACTTA 2940
3941 TCAAGGCTCTGTTTGTGCTGCTGCGCCATCCCGAAGCTATTTCAAGTACAGGCGCCAG 3000
3942 TCAAGGCTCTGTTTGTGCTGCTGCGCCATCCCGAAGCTATTTCAAGTACAGGCGCCAG 3000
3001 AGTCCCGAGAGATGTTCCAGATCTGTTTCAATCCGATGCTAGTTCACAAAGTGCAGGT 3060
3002 AGTCCCGAGAGATGTTTCCAGATCTGTTTCAATCCGATGCTAGTTCACAAAGTGCAGGT 3060
3061 TTTTGAGACCTTACAATGACTCAGCCACGCTGCTCAATCAAAATCTTCTGCTATAG 3120
3062 TTTTGAGACCTTACAATGACTCAGCCACGCTGCTCAATCAAAATCTTCTGCTATAG 3120
3121 GGTGTGGACAGAGCTGTCTTCTTCTGCTGATGTCAGCACAGTGGGTATTCCTGCTC 3180
3122 GGTGTGGACAGAGCTGTCTTCTTCTGCTGATGTCAGCACAGTGGGTATTCCTGCTC 3180
3181 CCGTATCAGTACTCATAGAGTTCAATTTTATATATATACAGATATTTTGGTAAAT 3240
3182 CCGTATCAGTACTCATAGAGTTCAATTTTATATATATACAGATATTTTGGTAAAT 3240
3241 GAACCTGGTCTTCTTCCAGCATCGTGATGAGCTGAGATGGCTTTGAGTGGCAT 3300
3242 GAACCTGGTCTTCTTCCAGCATCGTGATGAGCTGAGATGGCTTTGAGTGGCAT 3300
3301 CAGCTTCTCAGCTGTGGGGAGTCTTGGATAGATCAGCGCTGCTGAGTGGACT 3360
3302 CAGCTTCTCAGCTGTGGGGAGTCTTGGATAGATCAGCGCTGCTGAGTGGACT 3360
3361 TTGCTCAGCTAGTGTGAGACTCAGCTGCTCTCTGGGCTCTTACTCCTCCTCAAGGAGTC 3420
3362 TTGCTCAGCTAGTGTGAGACTCAGCTGCTCTCTGGGCTCTTACTCCTCCTCAAGGAGTC 3420
3421 TGTAGTGGAAAGAGGCGCCACAGAAATAGTGTCTTATCTGAAACTTCAGCTTCTCTTAGC 3480

3422 TGTAGTGGAAAGAGGCGCCACAGAAATAGCTCTTATCTGAAACTTCAGCTTCTCTAGC 3480
3481 CCGGCTCTCTTAAGGAGCTCTTGCAGCTGTGTGCAAGGCTCTGACAGGAGAAAGC 3540
3482 CCGGCTCTCTTAAGGAGCTCTTGCAGCTGTGTGCAAGGCTCTGACAGGAGAAAGC 3540
3541 CAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3600
3542 CAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3600
3601 ACTCAGTCTCTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3660
3602 ACTCAGTCTCTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3660
3661 AGTCTCAAGCAGTGTCTGCGGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3720
3662 AGTCTCAAGCAGTGTCTGCGGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3720
3721 AGCAGTCTTGGAGAGAT 3737
3722 AGCAGTCTTGGAGAGAT 3737

RESULT 2

US-10-177-293-44
; Sequence 44, Application US/10:77293
; Publication No. US20030124128A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Glatt, Karen
; APPLICANT: Zhao, Xumei
; APPLICANT: Ganraivaipu, Manjula
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Westens, Maureen
; APPLICANT: Wyer, Vic
; APPLICANT: Wang, Youzhen
; APPLICANT: Xu, Yongyao
; APPLICANT: Hoersch, Sebastian
; APPLICANT: Monahan, John
; APPLICANT: Meyers, Rachel E.
; APPLICANT: Bast Jr., Robert C.
; APPLICANT: Hortobagyi, Gabriel N.
; APPLICANT: Pasztai, Lajos
; APPLICANT: Meric, Funda
; APPLICANT: Sahin, Aysegül
; APPLICANT: Mills, Gordon B.
; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT,
; FILE REFERENCE: MRI-018
; CURRENT APPLICATION NUMBER: US/10:177,293
; PRIORITY FILING DATE: 2002-06-21
; PRIOR FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 60/325,887
; PRIOR FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 60/321,572
; PRIOR FILING DATE: 2001-06-27
; PRIOR APPLICATION NUMBER: US 60/326,501
; PRIOR FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: US 60/325,002
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 60/362,595
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 60/xxx,xxx
; PRIOR FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 506
; SOFTWARE: FastSeq for Windows version 4.0
; SEQ ID NO 44
; LENGTH: 3737
; TYPE: DNA
; ORGANISM: Homo sapiens

Query Match: 100.0% Score 3737; DB 14; Length: 3737;
US-10-177-293-44


```
QY 2161 GTGAGGCTGTGTCAACCTGGTGAATATTTCTGAGATGGCTTTGCACCTTGCACACTCT 2220
DB 2161 GTGAGGCTGTGTCAACCTGGTGAATATTTCTGAGATGGCTTTGCACCTTGCACACTCT 2220
QY 2221 GTGCCCTGGGCACTTCCAGCTGAAGTGTGCAACTTCTGCTTCCCTCTGTGGAGAG 2280
DB 2221 GTGCCCTGGGCACTTCCAGCTGAAGTGTGCAACTTCTGCTTCCCTCTGTGGAGAG 2280
QY 2281 GCCTTGCACCAACATCAGGAGCTACTTCCCTTTCAGGACTGTGAACACAGAGTTCAAT 2340
DB 2281 GCCTTGCACCAACATCAGGAGCTACTTCCCTTTCAGGACTGTGAACACAGAGTTCAAT 2340
QY 2341 GTTCACCTGGACATTTCTACAAACACCACTACCCGATGTATTCTGTGCCCAGTGGAA 2400
DB 2341 GTTCACCTGGACATTTCTACAAACACCACTACCCGATGTATTCTGTGCCCAGTGGAA 2400
QY 2401 CATACAGCTGAATTTGGAAAAAATAATTGTGTGTTTTCTTGCCAGGAAATACTACGACTG 2460
DB 2401 CATACAGCTGAATTTGGAAAAAATAATTGTGTGTTTTCTTGCCAGGAAATACTACGACTG 2460
QY 2461 ACTTTGATGGCTCCACAAACATACCCAGTGTAAACACAGAGATGTGGAGGGAGCTGG 2520
DB 2461 ACTTTGATGGCTCCACAAACATACCCAGTGTAAACACAGAGATGTGGAGGGAGCTGG 2520
QY 2521 GAGATTTTCACTGGGTACATTTGAATCCCAAACTACCCAGGCAATTTACCCAGCAACACCG 2580
DB 2521 GAGATTTTCACTGGGTACATTTGAATCCCAAACTACCCAGGCAATTTACCCAGCAACACCG 2580
QY 2581 AGTGTAGGTGACCATCAACCCAGCCCAAGCCGCGATCTGCTGATCGTGTCCCTGAGA 2640
DB 2581 AGTGTAGGTGACCATCAACCCAGCCCAAGCCGCGATCTGCTGATCGTGTCCCTGAGA 2640
QY 2641 TCTTCTGCCCATAGAGACGACTGTGGGACTATCTGGTGATCGGGAACCTCTTCAT 2700
DB 2641 TCTTCTGCCCATAGAGACGACTGTGGGACTATCTGGTGATCGGGAACCTCTTCAT 2700
QY 2701 CCAATTTCTGTGACACATATGAACCTGCCAGACTACGAACGCCCATCGCCCTTCACT 2760
DB 2701 CCAATTTCTGTGACACATATGAACCTGCCAGACTACGAACGCCCATCGCCCTTCACT 2760
QY 2761 CCAGGTCAAGAGCTGTGGATTGATTCAGTCAAGTCAATGAAGGAACAGCGCTAGAGGT 2820
DB 2761 CCAGGTCAAGAGCTGTGGATTGATTCAGTCAAGTCAATGAAGGAACAGCGCTAGAGGT 2820
QY 2821 TCCAGGTCCCATACGTGCATATGATGAGGACTACAGGAATCATTTGAAGACATAGTTC 2880
DB 2821 TCCAGGTCCCATACGTGCATATGATGAGGACTACAGGAATCATTTGAAGACATAGTTC 2880
QY 2881 GAGATGGCAGGCTCTATGCATCTGAGAACCATCAGGAAATCTTTAAGSATAAGAAACTTA 2940
DB 2881 GAGATGGCAGGCTCTATGCATCTGAGAACCATCAGGAAATCTTTAAGSATAAGAAACTTA 2940
QY 2941 TCAGGCTCTGTTGATGCTCTGGCCCATCCGAGACTATTTCAAGTACACAGCCCAAG 3000
DB 2941 TCAGGCTCTGTTGATGCTCTGGCCCATCCGAGACTATTTCAAGTACACAGCCCAAG 3000
QY 3001 AGTCCGAGAGATGTTTCCAAGATGTTTCAATCCGATGCTACGTTCCAAAGTGTCCAGGT 3060
DB 3001 AGTCCGAGAGATGTTTCCAAGATGTTTCAATCCGATGCTACGTTCCAAAGTGTCCAGGT 3060
QY 3061 TTTTGAGACCTTACAAATGACTCAGCCCACTGCGCACTCAATACAAATGTTTCTGCTATAG 3120
DB 3061 TTTTGAGACCTTACAAATGACTCAGCCCACTGCGCACTCAATACAAATGTTTCTGCTATAG 3120
QY 3121 GGTGGTGGGACAGAGCTGTTCTTCTGCTGATGTACAGCAGTGGGTATTGCTGCTC 3180
DB 3121 GGTGGTGGGACAGAGCTGTTCTTCTGCTGATGTACAGCAGTGGGTATTGCTGCTC 3180
QY 3181 CGGTATCAGTGACTCATTAGAGTTCAATTTTATAGATAAATACAGATATTTTGGTAAAT 3240
DB 3181 CGGTATCAGTGACTCATTAGAGTTCAATTTTATAGATAAATACAGATATTTTGGTAAAT 3240
```

```
QY 3241 GAACCTTGGTTTTCTTTCCAGCACTGTGGAATGTAGACTGAGAAATGGCTTTGAGTGGCAT 3300
DB 3241 GAACCTTGGTTTTCTTTCCAGCACTGTGGAATGTAGACTGAGAAATGGCTTTGAGTGGCAT 3300
QY 3301 CAGCTTCTCACTGCTGTGGGGGAGTCTCTTGGATAGATCAGGGCTGGCTGAGCTGGACT 3360
DB 3301 CAGCTTCTCACTGCTGTGGGGGAGTCTCTTGGATAGATCAGGGCTGGCTGAGCTGGACT 3360
QY 3361 TTGGTCAGCTAGGTGAGACTCACCTGTCTCTGTGGGCTTACTCTCTCTCTCAAGGAGTC 3420
DB 3361 TTGGTCAGCTAGGTGAGACTCACCTGTCTCTGTGGGCTTACTCTCTCTCTCAAGGAGTC 3420
QY 3421 TGTAGTGGAAAGGAGGCCACAGAAATAGCTCTTATCTGSAATCTTACGCTTCTCTTACG 3480
DB 3421 TGTAGTGGAAAGGAGGCCACAGAAATAGCTCTTATCTGSAATCTTACGCTTCTCTTACG 3480
QY 3481 CCGGCCCTCTCTAAGGGAGCCCTCTGCACTGCTGTGTGAGGCTCTGACCGAGCAGACAGG 3540
DB 3481 CCGGCCCTCTCTAAGGGAGCCCTCTGCACTGCTGTGTGAGGCTCTGACCGAGCAGACAGG 3540
QY 3541 CAAGAGGGAGGAAAGAGAGACCTTGCAAGGCTCCCTCCACCCACCTTGAGACCTGGGAGG 3600
DB 3541 CAAGAGGGAGGAAAGAGAGACCTTGCAAGGCTCCCTCCACCCACCTTGAGACCTGGGAGG 3600
QY 3601 ACTCAGTTTCTCCACAGCCTTCTCCAGCCTGTGTGATACAAAGTTTGATCCCGAGGAATTG 3660
DB 3601 ACTCAGTTTCTCCACAGCCTTCTCCAGCCTGTGTGATACAAAGTTTGATCCCGAGGAATTG 3660
QY 3661 AGTTCTAGCAGTGTCTGTGAATAAAGCAGAAAGCAAGAAATAGAAATAAATAAATACTA 3720
DB 3661 AGTTCTAGCAGTGTCTGTGAATAAAGCAGAAAGCAAGAAATAGAAATAAATAAATACTA 3720
QY 3721 AGCACTTCTGGAGACAT 3737
DB 3721 AGCACTTCTGGAGACAT 3737
```

RESULT 3

```
US-10-176-847-49
; Sequence 49, Application US/10176847
; Publication No. US20030069636A1
; GENERAL INFORMATION:
; APPLICANT: Velby, Petter Ole
; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF BREAST
; TITLE OF INVENTION: AND OVARIAN CANCER
; FILE REFERENCE: MRI-039
; CURRENT APPLICATION NUMBER: US/10/176,847
; CURRENT FILING DATE: 2002-06-21
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 49
; LENGTH: 3758
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-176-847-49
```

Query Match 98.3%; Score 3673.9; DB 14; Length 3758;

Best Local Similarity 99.8%; Pred No. 0;

Matches 3691; Conservative 0; Mismatches 2; Indels 6; Gaps 1;

```
QY 38 CCGCGCACTCCGCGCCTCTCCCGCAACCGCTGAGCCATCCATGCGGGTCCGCGGCG 97
DB 47 CCGCGCACTCCGCGCCTCTCCCGCAACCGCTGAGCCATCCATGCGGGTCCGCGGCG 106
QY 98 CAACCGTCCCGGGGCGGCTCTGCGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 157
DB 107 CAACCGTCCCGGGGCGGCTCTGCGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 160
QY 158 GATGCTGGGGGGGGCGGCTCCCGGGGCTGGGGCGGCTGCGGGGGCGCCGAGAGAGT 217
DB 161 GATGCTGGGGGGGGCGGCTCCCGGGGCTGGGGCGGCTGCGGGGGCGCCGAGAGAGT 220
```

QY 218 AGATGAGTGTGCCCAAGGGCTAGATGACTGCCATGCCGACGCCCTGTGTGACAAACACC 277
DB 221 AGATGAGTGTCCGCAAGGGCTAGATGACTGCCATGCCGACGCCCTGTGTGACAAACACC 280
QY 278 CACCTCCTCAAGAAGTGTCTGCAAGGCTGGCTACCAAGGGGAAGGCAGGAGTGTGAGGA 337
DB 281 CACCTCCTCAAGAAGTGTCTGCAAGGCTGGCTACCAAGGGGAAGGCAGGAGTGTGAGGA 340
QY 338 CATCGATGATGTGGAATGAGCTCAATGGAGGCTGTGTCCATGACACTGTTTGAATATTC 337
DB 341 CATCGATGATGTGGAATGAGCTCAATGGAGGCTGTGTCCATGACACTGTTTGAATATTC 450
QY 398 AGCAATATTCGTTGCCACTTGTGTTGATGGCTTCATGTTGGCTCATGACGGTCAATATTC 457
DB 401 AGCAATATTCGTTGCCACTTGTGTTGATGGCTTCATGTTGGCTCATGACGGTCAATATTC 460
QY 458 TCATTGATGTGACGAGTGTCTGCGAGAACAAATGGCGGTGCGAGCATACCTGTGTCAACGT 517
DB 461 TCATTGATGTGACGAGTGTCTGCGAGAACAAATGGCGGTGCGAGCATACCTGTGTCAACGT 520
QY 518 CATGGGAGCTATCAGTGTCTGCGAGAACAAATGGCGGTGCGAGCATACCTGTGTCAACGT 577
DB 521 CATGGGAGCTATCAGTGTCTGCGAGAACAAATGGCGGTGCGAGCATACCTGTGTCAACGT 580
QY 578 CTGCATTACCGCTCGGAAGAGGCTGAGCTGCATGAATAAGGATCACGCGCTGTAGTCA 637
DB 581 CTGCATTACCGCTCGGAAGAGGCTGAGCTGCATGAATAAGGATCACGCGCTGTAGTCA 640
QY 638 CATCTGMAAGAGGCCCAAGGGCAGGCTGCGCTGTGAGTGCAAGGCTGTTTGTAGCT 697
DB 641 CATCTGMAAGAGGCCCAAGGGCAGGCTGCGCTGTGAGTGCAAGGCTGTTTGTAGCT 750
QY 698 GGCCAAAGAACAGAGAGACTGCATCTTGACCTGTAAACATGGGAAACGGTGGTGCACGA 757
DB 702 GGCCAAAGAACAGAGAGACTGCATCTTGACCTGTAAACATGGGAAACGGTGGTGCACGA 760
QY 758 CTCTGTGACGATACAGCCGATGAGCCAGAGTGCAGCTGCCATCCACAGTACAAGATGCA 817
DB 761 CTCTGTGACGATACAGCCGATGAGCCAGAGTGCAGCTGCCATCCACAGTACAAGATGCA 820
QY 818 CACAGATGGGAGGAGCTGCTTTGAGTCAGAGGACACTGTCTGGAGTGCACAGAGCAA 977
DB 821 CACAGATGGGAGGAGCTGCTTTGAGTCAGAGGACACTGTCTGGAGTGCACAGAGCAA 980
QY 878 CACCACATCAGTGTGATGGGATAAACGGGTGAAACGGCGCTGTCTCAAGAAACGTG 937
DB 881 CACCACATCAGTGTGATGGGATAAACGGGTGAAACGGCGCTGTCTCAAGAAACGTG 940
QY 938 TGCTGTCAACAAATGGAGGCTGTGACCGACCTGTGAAGGATACCTTCGACAGGTGTCCA 997
DB 941 TGCTGTCAACAAATGGAGGCTGTGACCGACCTGTGAAGGATACCTTCGACAGGTGTCCA 1000
QY 998 CAGTTGCTCTTGTGATTCACCTCCAGTTGGATGGGAAGACATGTAAAGATATGATGA 1057
DB 1001 CAGTTGCTCTTGTGATTCACCTCCAGTTGGATGGGAAGACATGTAAAGATATGATGA 1060
QY 1058 GTGCCAGACCCGCAATGGAGGCTGTGATTCATTTCTGCAAAAACATCGTGGCGAGTTTGA 1117
DB 1061 GTGCCAGACCCGCAATGGAGGCTGTGATTCATTTCTGCAAAAACATCGTGGCGAGTTTGA 1120
QY 1118 CTGGCGCTGCAAGAAAGATTTAAATTAATAACAGATGAGAAGTCTTGCACAAATGTGGA 1177
DB 1121 CTGGCGCTGCAAGAAAGATTTAAATTAATAACAGATGAGAAGTCTTGCACAAATGTGGA 1180
QY 1178 TGAGTGTCTTTTGGATAGGACTGTGACACAGCTGCATCAACACCTTGGCAATTTGC 1237
DB 1181 TGAGTGTCTTTTGGATAGGACTGTGACACAGCTGCATCAACACCTTGGCAATTTGC 1240
QY 1238 TTGTGCTTGAACCGAGGATACCTGTATGGCTTCAACCCACTGTGGAGACACCAATGA 1297
DB 1241 TTGTGCTTGAACCGAGGATACCTGTATGGCTTCAACCCACTGTGGAGACACCAATGA 1300
QY 1298 GTGCAGCATCAACACGAGGCTGTGACAGGCTGTGTGTGAACACAGTGGGCGACTATGA 1357

DB 1301 GTGCAGCATCAACACGAGGCTGTGACAGGCTGTGTGTGAACACAGTGGGCGACTATGA 1360
QY 1358 ATGCCAGTGTCCACCTTGGGTACAAGCTCACTGGAAATAAAAAGACTGTGTGTGAAGTGA 1417
DB 1361 ATGCCAGTGTCCACCTTGGGTACAAGCTCACTGGAAATAAAAAGACTGTGTGTGAAGTGA 1420
QY 1418 GGGGTCTCTGCCCAACAAAGTGTGTACCCCGGTGTGTCCCTGGACTCGGTAAAGTGGTGG 1477
DB 1421 GGGGTCTCTGCCCAACAAAGTGTGTACCCCGGTGTGTCCCTGGACTCGGTAAAGTGGTGG 1480
QY 1478 AGGAGACGGTGTCTCTCCTCAGATGTCACTTGGCATTCACCTCTCTTCCAGATGTCAACAC 1537
DB 1491 AGGAGACGGTGTCTCTCCTCAGATGTCACTTGGCATTCACCTCTCTTCCAGATGTCAACAC 1540
QY 1538 CATCAGGACCAAGTGTAACTTTAAGCTTAAAGGCAAGTGTAGTTTGAATAAATGCTGA 1597
DB 1541 CATCAGGACCAAGTGTAACTTTAAGCTTAAAGGCAAGTGTAGTTTGAATAAATGCTGA 1600
QY 1598 GCTGTCTCTCCGAGGGTGTGCGACAGCACTACCAAGAGGACACAGTCAAGTAAAGAGAG 1657
DB 1601 GCTGTCTCTCCGAGGGTGTGCGACAGCACTACCAAGAGGACACAGTCAAGTAAAGAGAG 1660
QY 1658 CTCTCGCTACGTAAACCTTACATGACAGCTTGGCAAGCAAGTCCAGAGGAGCCCTTGGCG 1717
DB 1661 CTCTCGCTACGTAAACCTTACATGACAGCTTGGCAAGCAAGTCCAGAGGAGCCCTTGGCG 1720
QY 1718 ACCAAGACCCCTAAGGAATATGTTTATCACTGTTTGAAGTGTGAGCTTGAACCTTAAACAAA 1777
DB 1721 ACCAAGACCCCTAAGGAATATGTTTATCACTGTTTGAAGTGTGAGCTTGAACCTTAAACAAA 1780
QY 1728 GAGGTGTACAGCTTCTTGTGACCTGAGTGCATCTGAAGGCAACCGAGAGGCGCTCG 1837
DB 1781 GAGGTGTACAGCTTCTTGTGACCTGAGTGCATCTGAAGGCAACCGAGAGGCGCTCG 1840
QY 1838 TAAAGCCATCCCAACGCTCAGAAAGCGCTCCACAGGAGGAGTTCACCTCCAGCTCTC 1897
DB 1841 TAAAGCCATCCCAACGCTCAGAAAGCGCTCCACAGGAGGAGTTCACCTCCAGCTCTC 1900
QY 1898 AGGCATGAACCTTCGACGTGGCTTAAAGGCTTCCGACACATCTGAAGCGGAGAGTGC 1957
DB 1901 AGGCATGAACCTTCGACGTGGCTTAAAGGCTTCCGACACATCTGAAGCGGAGAGTGC 1960
QY 1958 CTGTGAGTGGCGCCAGGCTCATGCAAGAAACCAATGTGTGCTGAGTTCAGGCGTGGACCTTA 2017
DB 1961 CTGTGAGTGGCGCCAGGCTCATGCAAGAAACCAATGTGTGCTGAGTTCAGGCGTGGACCTTA 2020
QY 2018 TTATGATGGAGCAACGAAACGCTGCTATTTATGTCCAAATGGAACCTTCCAAAATGAGGA 2077
DB 2021 TTATGATGGAGCAACGAAACGCTGCTATTTATGTCCAAATGGAACCTTCCAAAATGAGGA 2080
QY 2078 AGGACAAATGACTGTGAACCATGCCAAGACCCAGAGAAATCTGGGGCTTGAAGACCCC 2137
DB 2081 AGGACAAATGACTGTGAACCATGCCAAGACCCAGAGAAATCTGGGGCTTGAAGACCCC 2140
QY 2138 AGAAGCTTGGAAATATCTGTAATGTGGAGGCTGTGTCAACCTGTGTGAATATCTTCGAGA 2197
DB 2141 AGAAGCTTGGAAATATCTGTAATGTGGAGGCTGTGTCAACCTGTGTGAATATCTTCGAGA 2200
QY 2198 TGGCTTTCACCTTCCAGCTCTGTGCCCTGGGACGTTCCAGCTTGAAGCTGTGTCGAAC 2257
DB 2201 TGGCTTTCACCTTCCAGCTCTGTGCCCTGGGACGTTCCAGCTTGAAGCTGTGTCGAAC 2260
QY 2258 TTCTGCTTCCCTGTGGAGGAGGCTTGGCAACCAACATCAGGAGGCTACTTCTTTCA 2317
DB 2261 TTCTGCTTCCCTGTGGAGGAGGCTTGGCAACCAACATCAGGAGGCTACTTCTTTCA 2320
QY 2318 GGAAGTGTGAACCAAGTGTCAATGTTCACCTGGACATTTCTACACACACACACCTTCAACG 2377
DB 2321 GGAAGTGTGAACCAAGTGTCAATGTTCACCTGGACATTTCTACACACACACACCTTCAACG 2380
QY 2378 ATGTATTCTTCTGCCCAGTGGGAACAATACAGCTTGAATTTGGAAAAATAATTTGTGTTTC 2437

Db 401 AGGCAATTATCGTTGCACTGTGTTTGTATGGCTCTCATTTGGCTCATGACGGTCAATAATTG 460
QY 458 TCTTGATGTGGACGAGTGTCTGGAGAACAAATGGGGCTTGGCAGCATACCTGTGTCAACGTT 517
Db 461 TCTTGATGTGGACGAGTGTCTGGAGAACAAATGGGGCTTGGCAGCATACCTGTGTCAACGTT 520
QY 518 CATGGGAGCTATGAGTGTCTGTGCAAGGAGGGGTTTTTCTGAGTGACAAATCAGACAC 577
Db 521 CATGGGAGCTATGAGTGTCTGTGCAAGGAGGGGTTTTTCTGAGTGACAAATCAGACAC 580
QY 578 CTGCAATTACCCCTCGGAAGAGGGCTGTAGCTGTGCAATGAATAAGGATCACGGCTGTAGTCA 637
Db 581 CTGCAATTACCCCTCGGAAGAGGGCTGTAGCTGTGCAATGAATAAGGATCACGGCTGTAGTCA 640
QY 638 CATCTGCAAGAGGCCCAAGGGGAGGGCTGTGCTGTGAGTGACAGGGCTGTGTTTTGAGCT 697
Db 641 CATCTGCAAGAGGCCCAAGGGGAGGGCTGTGCTGTGAGTGACAGGGCTGTGTTTTGAGCT 700
QY 698 GGCCAAAGAACAGAGAGACTGTGATCTTGACCTGTAAACCATGGGAACGGTGGGTGCCAGCA 757
Db 701 GGCCAAAGAACAGAGAGACTGTGATCTTGACCTGTAAACCATGGGAACGGTGGGTGCCAGCA 760
QY 758 CTCCTGTGACGATACAGCCGATGGCCAGAGTGCAGCTGCCATCCACAGTACAAAGATSCA 817
Db 761 CTCCTGTGACGATACAGCCGATGGCCAGAGTGCAGCTGCCATCCACAGTACAAAGATSCA 820
QY 818 CACAGATGGGAGAGTGCCTTGGGGAGAGAGACACTGTCTGAGAGGTGACAGAGACAA 877
Db 821 CACAGATGGGAGAGTGCCTTGGGGAGAGAGACACTGTCTGAGAGGTGACAGAGACAA 880
QY 878 CACCACATCAGTGGTGGATGGGGATAAAGCGGCTGAAACGGGGCTGTCTCATGGAAACGTTG 937
Db 881 CACCACATCAGTGGTGGATGGGGATAAAGCGGCTGAAACGGGGCTGTCTCATGGAAACGTTG 940
QY 938 TGCCTGTCAACATGAGAGCTGTGACCGACACCTGTGAAGATTAATTCGACAGGTGTCCACTG 997
Db 941 TGCCTGTCAACATGAGAGCTGTGACCGACACCTGTGAAGATTAATTCGACAGGTGTCCACTG 1000
QY 998 CAGTGTCTCTGTGATTCACCTCTCCAGTTGATGGGAAGACATGTAAAGATATTGATGA 1057
Db 1001 CAGTGTCTCTGTGATTCACCTCTCCAGTTGATGGGAAGACATGTAAAGATATTGATGA 1060
QY 1058 GTGCCAGACCCGCAATGGAGTGTGTAATCATTTTCGCAAAACATCGTGGCAGTTTTGA 1117
Db 1061 GTGCCAGACCCGCAATGGAGTGTGATCATTTTCGCAAAACATCGTGGCAGTTTTGA 1120
QY 1118 CTGGGCTGCAAGAAGAGATTTAAATTAATTAACAGATCAGAAAGTCTTCCCAAGATGTGA 1177
Db 1121 CTGGGCTGCAAGAAGAGATTTAAATTAATTAACAGATCAGAAAGTCTTCCCAAGATGTGA 1180
QY 1178 TGAGTGTCTTTTGGATAGGACTGTGACCAAGCTGTGCATCAACCAACCTGGCACATTTGC 1237
Db 1181 TGAGTGTCTTTTGGATAGGACTGTGACCAAGCTGTGCATCAACCAACCTGGCACATTTGC 1240
QY 1238 TTGTGCTTGCACCGAGGTACACCTGTATGGCTTACCCACTGTGGAGACCCATGA 1297
Db 1241 TTGTGCTTGCACCGAGGTACACCTGTATGGCTTACCCACTGTGGAGACCCATGA 1300
QY 1298 GTGCAGCATCAACAGGAGGCTGTGAGCAGAGTCTGTGTGAACACAGTGGGACGCTATGA 1357
Db 1301 GTGCAGCATCAACAGGAGGCTGTGAGCAGAGTCTGTGTGAACACAGTGGGACGCTATGA 1360
QY 1358 ATGCCAGTGGCACCCTGGGTACAGCTCCACTGGGAATTAATAAGACTGTGTGGAAGTGA 1417
Db 1361 ATGCCAGTGGCACCCTGGGTACAGCTCCACTGGGAATTAATAAGACTGTGTGGAAGTGA 1420
QY 1418 GGGGCTCTGCCACAAAGTGTGTACCCCGTGTGTCCCTGCACTGGCGTAGAGGTGGTGG 1477
Db 1421 GGGGCTCTGCCACAAAGTGTGTACCCCGTGTGTCCCTGCACTGGCGTAGAGGTGGTGG 1480
QY 1478 AGGAGACGGGTGCTTCTCAGATGTCACTGTGGCATTTCACTCTCTTCCAGATGTCAACAC 1537
Db 1481 AGGAGACGGGTGCTTCTCAGATGTCACTGTGGCATTTCACTCTCTTCCAGATGTCAACAC 1540

QY 1538 CATCAGGACAAAGTGTAAACCTTTAAAGCTAAATGAAGGCAAGTGTAGTTTGAAAAATGCTGA 1597
Db 1541 CATCAGGACAAAGTGTAAACCTTTAAAGCTAAATGAAGGCAAGTGTAGTTTGAAAAATGCTGA 1600
QY 1598 GCTGTTTCCGAGGGTCTGGACCCAGCACTACAGAGAACACAGCTCACTAATAAGAGAG 1657
Db 1601 GCTGTTTCCGAGGGTCTGGACCCAGCACTACAGAGAACACAGCTCACTAATAAGAGAG 1660
QY 1658 CTTCCGCTACGCTAAACCTTACATGAGCTGTGSSAAGCAAGTCCCAGGAGCCCTGSCCG 1717
Db 1661 CTTCCGCTACGCTAAACCTTACATGAGCTGTGSSAAGCAAGTCCCAGGAGCCCTGSCCG 1720
QY 1718 ACCAAGCACCCCTAAGGAAATGTTTATCACTGTGTTGAGTTGAGCTTGAATAACCAAAA 1777
Db 1721 ACCAAGCACCCCTAAGGAAATGTTTATCACTGTGTTGAGTTGAGCTTGAATAACCAAAA 1780
QY 1778 GGAGGTGACAGCTTCTTTGTGACCTGAGCTGCATCGTAAAGCGAACCGAGAGCGGCTCCG 1837
Db 1781 GGAGGTGACAGCTTCTTTGTGACCTGAGCTGCATCGTAAAGCGAACCGAGAGCGGCTCCG 1840
QY 1838 TAAAGCCATCCGACAGCTCAGAAAGGGCTTCCAGGGAGCAGTTTCACTCCAGCTCTC 1897
Db 1841 TAAAGCCATCCGACAGCTCAGAAAGGGCTTCCAGGGAGCAGTTTCACTCCAGCTCTC 1900
QY 1898 AGGCTGGAACCTCGACAGCTGTGCTAAAGAGCCCTCCAGACATCTGAAGCCGACAGAGTC 1957
Db 1901 AGGCTGGAACCTCGACAGCTGTGCTAAAGAGCCCTCCAGACATCTGAAGCCGACAGAGTC 1960
QY 1958 CTGTGGATGGGCCACAGSCTCATGCAAGAAACCAATGTGTGAGTTGCAAGGCTGGAGACTA 2017
Db 1961 CTGTGGATGGGCCACAGSCTCATGCAAGAAACCAATGTGTGAGTTGCAAGGCTGGAGACTA 2020
QY 2018 TTATGATGGAGCAGAGAACCTGCTGATTTATGTCCAAATGGAACTTCCAAATGAGGA 2077
Db 2021 TTATGATGGAGCAGAGAACCTGCTGATTTATGTCCAAATGGAACTTCCAAATGAGGA 2080
QY 2078 AGGCAAAATGACTTGTGAACCATGCCAAGACACAGGAAATTCGGGGCCCTGAAGACCCC 2137
Db 2081 AGGCAAAATGACTTGTGAACCATGCCAAGACACAGGAAATTCGGGGCCCTGAAGACCCC 2140
QY 2138 AGAAGCTTGGAAATGTCTGAATGTGAGGTCTGTGTCAACCTGGTGAATTTCTGAGA 2197
Db 2141 AGAAGCTTGGAAATGTCTGAATGTGAGGTCTGTGTCAACCTGGTGAATTTCTGAGA 2200
QY 2198 TGGCTTTGCACTTGGCAGCTCTGTCCTGGGACCTTCCAGCCTGAAGCTGGTCCAAAC 2257
Db 2201 TGGCTTTGCACTTGGCAGCTCTGTCCTGGGACCTTCCAGCCTGAAGCTGGTCCAAAC 2260
QY 2258 TTCTGTCTTCCCTGTGGAGGAGGCTTGCACCAAAACA-CAGGAGCTACTTCTCTTCA 2317
Db 2261 TTCTGTCTTCCCTGTGGAGGAGGCTTGCACCAAAACA-CAGGAGCTACTTCTCTTCA 2320
QY 2318 GAGCTGTCAAAACAGAGTCAAACTGTTCACCTGGACATTTCTAGAACACCACTACCC 2377
Db 2321 GAGCTGTCAAAACAGAGTCAAACTGTTCACCTGGACATTTCTAGAACACCACTACCC 2380
QY 2378 ATGTATTCTGTTGCCAGTGGGAACATACCAAGCTTGAATTTGGAAAAAATTAATTTGTTTC 2437
Db 2381 ATGTATTCTGTTGCCAGTGGGAACATACCAAGCTTGAATTTGGAAAAAATTAATTTGTTTC 2440
QY 2438 TTGCCAGGAATACTACGACTGACTTGTGATGGTCCAGAAATTAACCCAGTGTAAAAA 2497
Db 2441 TTGCCAGGAATACTACGACTGACTTGTGATGGTCCAGAAATTAACCCAGTGTAAAAA 2500
QY 2498 CAGAAGATGTGGAGGAGTGGCAGATTTCACTGGGTACATTTGAATCCCAAACTACCC 2557
Db 2501 CAGAAGATGTGGAGGAGTGGCAGATTTCACTGGGTACATTTGAATCCCAAACTACCC 2560
QY 2558 AGGCAATTTACCAAGCCCAACAGCGAGTGTAGTGAACCATCAACCCCAAGGAGGCGG 2617
Db 2561 AGGCAATTTACCAAGCCCAACAGCGAGTGTAGTGAACCATCAACCCCAAGGAGGCGG 2620

QY 2618 CATCTGATCGTGGTCCCTGAGATCTTCTGCCCCATAGAGACGACTGTGGGACTATCT 2677
 Db 2621 CATCTGATCGTGGTCCCTGAGATCTTCTGCCCCATAGAGACGACTGTGGGACTATCT 2680
 QY 2678 GGTGATCGGGAACCTCTTTCATCAATTCGTGACACATATGAACCTCCAGACCTA 2737
 Db 2691 GGTGATCGGGAACCTCTTTCATCAATTCGTGACACATATGAACCTCCAGACCTA 2740
 QY 2738 CGAAGCCCCCATCGCTTCACTCCAGCTCAAGAGCTGTGGATTGAGTTCAAGTCCAA 2797
 Db 2741 CGAAGCCCCCATCGCTTCACTCCAGCTCAAGAGCTGTGGATTGAGTTCAAGTCCAA 2800
 QY 2798 TGAAGGAACAGCGCTAGAGGTTCCAGGTCCCATACGTGACATATGATGAGGACTACCA 2857
 Db 2801 TGAAGGAACAGCGCTAGAGGTTCCAGGTCCCATACGTGACATATGATGAGGACTACCA 2860
 QY 2858 GGAACCTATTGAAGACATAGTTCGAGATGGAGGCTCTATCACTGAGAACCATCAGGA 2917
 Db 2861 GGAACCTATTGAAGACATAGTTCGAGATGGAGGCTCTATCACTGAGAACCATCAGGA 2920
 QY 2918 AATACTTAAGGATAGAAACTTATCAAGGCTCTGTTGATGTCCTGGGCCATCCCCAGAA 2977
 Db 2921 AATACTTAAGGATAGAAACTTATCAAGGCTCTGTTGATGTCCTGGGCCATCCCCAGAA 2980
 QY 2978 CTATTTCAAGTACACAGCCAGGATCCGAGAGATGTTTCCAGATCGTTTCATCCGATT 3037
 Db 2981 CTATTTCAAGTACACAGCCAGGATCCGAGAGATGTTTCCAGATCGTTTCATCCGATT 3040
 QY 3038 GCTAGCTTCCAAAGTGTCCAGGTTTTGAGACCTTACAAATGACTCAGCCACCTGGCCAC 3097
 Db 3041 GCTAGCTTCCAAAGTGTCCAGGTTTTGAGACCTTACAAATGACTCAGCCACCTGGCCAC 3100
 QY 3098 TCAATACAAATGTTCTGTATAGGTTTGGTGGACAGAGCTGTCTTCTCTGATGTCA 3157
 Db 3101 TCAATACAAATGTTCTGTATAGGTTTGGTGGACAGAGCTGTCTTCTCTGATGTCA 3160
 QY 3158 CCACAGTGGGATGTTGCTGCTCCCGTATCAGTGACTATTAGATTTCAATTTTATAGA 3217
 Db 3161 CCACAGTGGGATGTTGCTGCTCCCGTATCAGTGACTATTAGATTTCAATTTTATAGA 3220
 QY 3218 TAATACAGATATTTTGGTAAATGAATGTTGTTTTTCTTCCAGCATCTGGGATGAGA 3277
 Db 3221 TAATACAGATATTTTGGTAAATGAATGTTGTTTTTCTTCCAGCATCTGGGATGAGA 3280
 QY 3278 CTGAGAATGGCTTTCAGTGGCATAGCTTCTACCTGCTGTGGGCGATGTCTTGGATAGA 3337
 Db 3281 CTGAGAATGGCTTTCAGTGGCATAGCTTCTACCTGCTGTGGGCGATGTCTTGGATAGA 3340
 QY 3338 TCACGGGCTGGCTGAGCTGGATTTGGTCAGCTAGGTGAGACTCACCCTGCTCTGGG 3397
 Db 3341 TCACGGGCTGGCTGAGCTGGATTTGGTCAGCTAGGTGAGACTCACCCTGCTCTGGG 3400
 QY 3398 GTCTTACTCTCTCTCAAGGAGTCTGTAGTGGAAAGAGGCCACAGATAAGTCTCTATT 3457
 Db 3401 GTCTTACTCTCTCTCAAGGAGTCTGTAGTGGAAAGAGGCCACAGATAAGTCTCTATT 3460
 QY 3458 CTGAAATCTCAGCTTCTCTAGCCGGGCTCTCTAAGGGAGCCCTCTGCACTGTGTGTC 3517
 Db 3461 CTGAAATCTCAGCTTCTCTAGCCGGGCTCTCTAAGGGAGCCCTCTGCACTGTGTGTC 3520
 QY 3518 AGGCTCTGACAGGAGACAGGAGAGGAGGAGGAGGAGCCCTGCAAGGCTCCCTC 3577
 Db 3521 AGGCTCTGACAGGAGACAGGAGAGGAGGAGGAGGAGGAGCCCTGCAAGGCTCCCTC 3580
 QY 3578 CACCCACCTTGAGACTGGGAGGACTCAGTTTCTCCACAGCCTTCTCCAGGCTGTGTGAT 3637
 Db 3581 CACCCACCTTGAGACTGGGAGGACTCAGTTTCTCCACAGCCTTCTCCAGGCTGTGTGAT 3640
 QY 3638 ACAAGTTTGAATCCAGGAATTTGAGTTCTTAAGAGTGTCTGTGAAAAAAGCAGAAA 3697
 Db 3641 ACAAGTTTGAATCCAGGAATTTGAGTTCTTAAGAGTGTCTGTGAAAAAAGCAGAAA 3700
 QY 3698 GAATTAGAAATAATAAAAAACTTAAGCACTTCTTGGAGACA 3736

Db 3701 GAATTAGAAATAATAAAAAACTTAAGCACTTCTTGGAGACA 3739
 RESULT 5
 US-10-198-846-11192
 : Sequence 1192, Application US/0198846
 : Publication No. US2003099974A1
 : GENERAL INFORMATION:
 : APPLICANT: Xu, Yongyao
 : APPLICANT: Wang, Youzhen
 : APPLICANT: Steelmann, Kathleen
 : TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
 : TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
 : TITLE OF INVENTION: THERAPY OF BREAST CANCER
 : FILE REFERENCE: MRI-C49
 : CURRENT APPLICATION NUMBER: US/10/198,846
 : CURRENT FILING DATE: 2002-07-18
 : PRIOR APPLICATION NUMBER: 60/306,220
 : PRIOR FILING DATE: 2001-07-18
 : NUMBER OF SEQ ID NOS: 14084
 : SOFTWARE: FastSeq for Windows Version 4.0
 : SEQ ID NO 11192
 : LENGTH: 3091
 : TYPE: DNA
 : ORGANISM: Homo sapiens
 : FEATURE:
 : NAME/KEY: misc feature
 : LOCATION: 1..27978, 2973, 2980, 2981, 2982, 2983, 3085, 3086, 3087,
 : LOCATION: 3088, 3089, 3090, 3091
 : OTHER INFORMATION: n = A, T, C or G
 US-10-198-846-11192
 Query Match: 74.0%; Score 2763.6; DB 14; Length 3091;
 Best Local Similarity 96.9%; Pred. No. 0;
 Matches 2863; Conservative 0; Mismatches 4; Indels 87; Gaps 1;
 QY 377 CCAATGACTGTTTGAATATTCCAGGCAATTATCGTGGCACTTTTGTGATGGCTTATGTT 436
 Db 16 CGATGACTGTTTGAATATTCCAGGCAATTATCGTGGCACTTTTGTGATGGCTTATGTT 75
 QY 437 GGCTCATGAGCGTCATCAATGCTTGTGATGTCAGAGTGCCTGGAGAACTATGGGGCTG 496
 Db 76 GGCTCATGAGCGTCATCAATGCTTGTGATGTCAGAGTGCCTGGAGAACTATGGGGCTG 135
 QY 497 CCAGCATACCTGTGTCACGCTCATGCGGAGCTATGAGTCTGCTGCAAGAGGGGTTTTT 556
 Db 136 CCAGCATACCTGTGTCACGCTCATGCGGAGCTATGAGTCTGCTGCAAGAGGGGTTTTT 195
 QY 557 CCTGAGTGCATATCAGCACACCTGCAATTCACCGCTCGAAGAGGGGCTGAGCTGCATGAA 616
 Db 196 CCTGAGTGCATATCAGCACACCTGCAATTCACCGCTCGAAGAGGGGCTGAGCTGCATGAA 255
 QY 617 TAAGGATCAGCGCTGTAGTCATCTGCAAGAGGGGCCCCAAGGGGAGCGTGCCTGTGA 676
 Db 256 TAAGGATCAGCGCTGTAGTCATCTGCAAGAGGGGCCCCAAGGGGAGCGTGCCTGTGA 315
 QY 677 GTCCAGCGCTGTTTGTGAGCTGCCCAAGAACACAGAGAGACTGCATCTTGACCTGTAAACA 736
 Db 316 GTCCAGCGCTGTTTGTGAGCTGCCCAAGAACACAGAGAGACTGCATCTTGACCTGTAAACA 375
 QY 737 TGGAAACGTTGGTGGCCAGCACTCTCTGTGACGATACAGCCGATGGCCCGAGAGTGCAGCTG 796
 Db 376 TGGAAACGTTGGTGGCCAGCACTCTCTGTGACGATACAGCCGATGGCCCGAGAGTGCAGCTG 435
 QY 797 CCATCCAGCTACAGATGCACACATGCGAGGAGCTGCCCTGAGCGAGAGAGACTGCT 856
 Db 436 CCATCCAGCTACAGATGCACACATGCGAGGAGAGCTGCCCTGAGCGAGAGAGACTGCT 495
 QY 857 CTTGGAGGTGACAGAGAGCAACACCATCATCAGTGGTGGATGGGATTAACCGGTGAACG 916
 Db 496 CTTGGAGGTGACAGAGAGCAACACCATCATCAGTGGTGGATGGGATTAACCGGTGAACG 555

QY	917	GCGGCTGCTCATGGAAACGTGTGCTGTCACAATGGAGGCTGTGACCGCACCC ⁻ GTGAAGGA	976
DB			
DB	556	GCGGCTGCTCATGGAAACGTGTGCTGTCACAATGGAGGCTG ⁻ GACCGCACCTGTGAAGGA	615
QY	977	TACTTGGAGAGTGTCGACTGCAGTTGTGCTGTTGSA ⁻ TTCACCTCTCCAGTTGATGGGAA	1036
DB			
DB	616	TACTTGCACAGGTGCCACTGCAGTTGTGCTGTTGSA ⁻ TTCACCTCTCCAGTTGATGGGAA	675
QY	1037	GACATGTAAGAATATTGATGAGTGCACAGACCCGCAATGGAGGTGTGATCAT ⁻ TTCGCCAA	1096
DB			
DB	676	GACATGTAAGAATATTGATGAGTGCACAGACCCGCAATGGAGGTGTGATCA ⁻ TTCGCCAA	735
QY	1097	AACATCGTGGCAGTTTTGACTGCGCTGCAAGAAAAGGATTTAAATATTATAACAGATGA	1156
DB			
DB	736	AACATCGTGGCAGTTTTGACTGCGCTGCAAGAAAAGGATTTAAATATTATAACAGATGA	795
QY	1157	GAGTCTTGCCAGATGCGATGATGCTCTTTGGTAGGACCTGTGACCAGAGCTGCAT	1216
DB			
DB	796	GAGTCTTGCCAGATGCGATGATGCTCTTTGGTAGGACCTGTGACCAGAGCTGCAT	855
QY	1217	CAACCACTCTGGCACATTTGCTTGTGCTTGCACCCGAGGGTACACCTGTATGGCTTCAC	1276
DB			
DB	856	CAACCACTCTGGCACATTTGCTTGTGCTTGCACCCGAGGGTACACCTGTATGGCTTCAC	915
QY	1277	CCACTGTGGAGACACCAATGAGTGCAGCATCAACAACGGAGGGTGTGACGAGGTC ⁻ GTGT	1336
DB			
DB	916	CCACTGTGGAGACACCAATGAGTGCAGCATCAACAACGGAGGGTGTGACGAGGTC ⁻ GTGT	975
QY	1337	GAACACAGTGGCAGCTATGAATGCCAGTSCACCC ⁻ GGGTACAAGCTCCACCTGGGAATAA	1396
DB			
DB	976	GAACACAGTGGCAGCTATGAATGCCAGTSCACCC ⁻ GGGTACAAGCTCCACCTGGGAATAA	1035
QY	1397	AAAAGACTGTGTGGAAGTGAAGGGCTCTCTGCCACAAGTGTGTCA ⁻ CCCCGTGTG ⁻ CCCT	1456
DB			
DB	1036	AAAAGACTGTGTGGAAGTGAAGGGCTCTCTGCCACAAGTGTGTCA ⁻ CCCCGTGTG ⁻ CCCT	1095
QY	1457	GCACTGGGTTAAGTGGTGAGAGACGGTGTCTCTCAGATGTCACTCTGGCATTCOA	1516
DB			
DB	1096	GCACTGGGTTAAGTGGTGAGAGACGGTGTCTCTCAGATGTCACTCTGGCATTCOA	1155
QY	1517	CCTCTCTTCA-----	1526
DB			
DB	1156	CCTCTCTCAGGACTGCAAGGGGCTACTGTGCACCTGTGGCTCTCTCTCTCTCTCAG	1215
QY	1527	-----GATGTCAACCACTACAGGACAAG	1549
DB			
DB	1216	GAACAAACAACAAAAATCAAATGACTGTCTTTTGGGGATGTCAACCACTACAGGACAAG	1275
QY	1550	TGTAACTTTTAAGCTAAATGAAGCAAGTGTAGTTTGA AAAAATGCTGAGCTGTTTCCC GA	1609
DB			
DB	1276	TGTAACTTTTAAGCTAAATGAAGCAAGTGTAGTTTGA AAAAATGCTGAGCTGTTTCCC GA	1335
QY	1610	GGGTCTGCGACCACTACACAGAGAAGCACAGCTCAGTAAAGAGAGCTTCCGCTACGT	1669
DB			
DB	1336	GGGTCTGCGACCACTACACAGAGAAGCACAGCTCAGTAAAGAGAGCTTCCGCTACGT	1395
QY	1670	AAACCTTACATGCACTCTGGCAGCAAGTCCCAAGGACCCCTGGCCGA ⁻ CCAAGCACCCC	1729
DB			
DB	1396	AAACCTTACATGCACTCTGGCAGCAAGTCCCAAGGACCCCTGGCCGA ⁻ CCAAGCACCCC	1455
QY	1730	TAAGAAATGTTTATCACTGTGAGTTTGAGCTTTGAAACTAACCAAAAGAGAGGTGACAGC	1789
DB			
DB	1456	TAAGAAATGTTTATCACTGTGAGTTTGAGCTTTGAAACTAACCAAAAGAGAGGTGACAGC	1515
QY	1790	TTCTTGTGACCTGAGCTGCATCGTAAAGCGAACCCGAGAAGCGGCTCCGTAAAGCCATCCG	1849
DB			
DB	1516	TTCTTGTGACCTGAGCTGCATCGTAAAGCGAACCCGAGAAGCGGCTCCGTAAAGCCATCCG	1575
QY	1850	CACGCTCAGAAAGCGCTCCACAGGGAGCAGTTTCACTCTCAGCTCTCAGGCATGAACCT	1909
DB			
DB	1576	CACGCTCAGAAAGCGCTCCACAGGGAGCAGTTTCACTCTCAGCTCTCAGGCATGAACCT	1635

QY	1910	CGACGTGGCTAAAGAGCCTCCGACAGACATCTGAA	CGCCAGGCGAGAGTCTCTCTGTGGAGCTGGG	1969
DB	1636	CGACGTGGGTAAAGAGCCTCCGACAGACATCTGAA	CGCCAGGCGAGAGTCTCTCTGTGGAGCTGGG	1695
QY	1970	CCAGGGTCATGCAGAAAAACCAATGTGTCAAGTTG	CCAGGGCTGGGACCCATATATGATGGAGC	2029
DB	1696	CCAGGGTCATGCAGAAAAACCAATGTGTCAAGTTG	CCAGGGCTGGGACCCATATATGATGGAGC	1755
QY	2030	ACGAGAACCGTGCATTTTATGTCCAAATGGAACTT	TCCAAATATGAGAAAGGACAAATGAC	2089
DB	1756	ACGAGAACCGTGCATTTTATGTCCAAATGGAACTT	TCCAAATATGAGAAAGGACAAATGAC	1815
QY	2090	TTGTGAAACGATGCCAAGACCCAGGAAATTTCT	TGGGGCCCTGAGAGCCCAAGAGAGCTTGGAA	2149
DB	1816	TTGTGAAACGATGCCAAGACCCAGGAAATTTCT	TGGGGCCCTGAGAGCCCAAGAGAGCTTGGAA	1875
QY	2150	TATGTCTGAATGTGGAGGTCCTGTGTCAACCTGGT	GTGAAATTTCTGCAGATGCGTTTGCACC	2209
DB	1876	TATGTCTGAATGTGGAGGTCCTGTGTCAACCTGGT	GTGAAATTTCTGCAGATGCGTTTGCACC	1935
QY	2210	TTGCCAGCTCTGTGSCCTCTGGGACGTTCCAGSCT	GTGAGAGCTGTGTGAACTTCTGCTTCCG	2269
DB	1936	TTGCCAGCTCTGTGSCCTCTGGGACGTTCCAGSCT	GTGAGAGCTGTGTGAACTTCTGCTTCCG	1995
QY	2270	CTGTGGAGGAGGCTTGGCCACCAAAACATCAGGAG	GCATTTCTCTTCCAGGACCTGTGAAC	2329
DB	1996	CTGTGGAGGAGGCTTGGCCACCAAAACATCAGGAG	GCATTTCTCTTCCAGGACCTGTGAAC	2055
QY	2330	CAGAGTTCAATGTCTACCTGTGGACATTTCTTACA	CACACCACTCAACCGATGTATTCGTTG	2389
DB	2056	CAGAGTTCAATGTCTACCTGTGGACATTTCTTACA	CACACCACTCAACCGATGTATTCGTTG	2115
QY	2390	CCCAGTGGGAACATACAGAGCTCAATTTGGA	AAAAAATAATTGTGTTCTTGGCCACGAAA	2449
DB	2116	CCCAGTGGGAACATACAGAGCTCAATTTGGA	AAAAAATAATTGTGTTCTTGGCCACGAAA	2175
QY	2450	TACTACGACTGACTTTGATGGGTCTCCACAAACAT	ATACCCAGTGTGAAAAACAGAAAGTGTGG	2509
DB	2176	TACTACGACTGACTTTGATGGGTCTCCACAAACAT	ATACCCAGTGTGAAAAACAGAAAGTGTGG	2235
QY	2510	AGSGGAGCTGGGAGATTTTCACTGGGTACACTTGA	ATCTCCCAAACTACCCAGGCGAATTACCC	2569
DB	2236	AGSGGAGCTGGGAGATTTTCACTGGGTACACTTGA	ATCTCCCAAACTACCCAGGCGAATTACCC	2295
QY	2570	AGSCAACACCGAGTGTACGTTGACCATCAACCCAC	CCGCGCGCATCTCTGTATCGT	2629
DB	2296	AGSCAACACCGAGTGTACGTTGACCATCAACCCAC	CCGCGCGCATCTCTGTATCGT	2355
QY	2630	GGTCCCTGAGACTTCTCTGCCCTAGAGACAGACT	GTGGGGAGCTCTCTGTGATCGGAA	2689
DB	2356	GGTCCCTGAGACTTCTCTGCCCTAGAGACAGACT	GTGGGGAGCTCTCTGTGATCGGAA	2415
QY	2690	AACCTCTTCATCCAAATTTCTGTGCAACACATGA	ATGAACCTGTCCAGACCTACGAACGCCCAT	2749
DB	2416	AACCTCTTCATCCAAATTTCTGTGCAACACATGA	ATGAACCTGTCCAGACCTACGAACGCCCAT	2475
QY	2750	CGSCTTCACCTCCAGSTCAAGAGAGCTGTGGATT	CAAGTTCAAGTCCAAATCAAGGGAGACAG	2809
DB	2476	CGSCTTCACCTCCAGSTCAAGAGAGCTGTGGATT	CAAGTTCAAGTCCAAATCAAGGGAGACAG	2535
QY	2810	CGGTAGAGGCTTCCAGGTCCTCATACGTGCACATG	ATGATGAGGACTACACAGGAACTCATTTGA	2869
DB	2536	CGGTAGAGGCTTCCAGGTCCTCATACGTGCACATG	ATGATGAGGACTACACAGGAACTCATTTGA	2595
QY	2870	AGACATAGTTCGAGATGGGACGGCTCTATGCATCT	GAGAACCATCAGGAAATACCTTAAGGA	2929
DB	2596	AGACATAGTTCGAGATGGGACGGCTCTATGCATCT	GAGAACCATCAGGAAATACCTTAAGGA	2655
QY	2930	TAGAAACCTTATCAAGGCTGTGTTTGAATGCTCT	TGCTTGGCCCATCCCAAGACATATTTCAAGTA	2989
DB	2656	TAGAAACCTTATCAAGGCTGTGTTTGAATGCTCT	TGCTTGGCCCATCCCAAGACATATTTCAAGTA	2715
QY	2990	CACAGTCCAGGAGTCCCGAGAGATGTTTTCAGAGAT	CTGTTCAATCGGATTTCTAGTCTCCAA	3049

DB 2726 CACAGCCAGGAGTCCCGAGAGATGTTCCAGATCGTTTCATCCGATGCTACGTTCCAA 2775
QY 3050 AGTGTCCAGGTTTTCAGACCTTTACAAATGACTCAGCCACCGTGCACCTCAATACAAATG 3109
DB 2776 AGTGTCCAGGTTTTCAGACCTTTACAAATGACTCAGCCACCGTGCACCTCAATACAAATG 2835
QY 3110 TTCTGCTATAGGTTTGGTGGGACAGAGCTGCTTCCTTCTGCTATGTCAGCACAGTCGGGT 3169
DB 2836 TTCTGCTATAGGTTTGGTGGGACAGAGCTGCTTCCTTCTGCTATGTCAGCACAGTCGGGT 2895
QY 3170 ATTGCTGCTCCCGTATCAGTCACTATTAGTAGTTCATTTTATAGATAATACAGATAT 3229
DB 2896 ATTGCTGCTCCCGTATCAGTCACTATTAGTAGTTCATTTTATAGATAATACAGATAT 2955
QY 3230 TTGTGTAATGAA 3243
DB 2956 TTGTGTAATGAA 2969

RESULT 6

US-09-898-570-11/c
; Sequence 11, Application US/09898570
; Patent No. US20020123612A1
; GENERAL INFORMATION:
; APPLICANT: GERLACH, VALERIE L.
; APPLICANT: ELLERMAN, KAREN
; APPLICANT: MACDOUGALL, JOHN R.
; APPLICANT: SMITHSON, GLENNDA
; TITLE OF INVENTION: NOVEL HUMAN PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND
; TITLE OF INVENTION: METHODS OF USING THE SAME
; FILE REFERENCE: 15966-776CIP

; CURRENT APPLICATION NUMBER: US/09/898,570
; CURRENT FILING DATE: 2001-07-03
; PRIOR APPLICATION NUMBER: 60/198,293
; PRIOR FILING DATE: 2000-04-19
; PRIOR APPLICATION NUMBER: 60/198,645
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: 60/210,809
; PRIOR FILING DATE: 2000-06-09
; PRIOR APPLICATION NUMBER: 60/199,476
; PRIOR FILING DATE: 2000-04-26
; PRIOR APPLICATION NUMBER: 60/200,025
; PRIOR FILING DATE: 2000-04-26
; PRIOR APPLICATION NUMBER: 60/224,610
; PRIOR FILING DATE: 2000-08-11
; PRIOR APPLICATION NUMBER: 60/200,024
; PRIOR FILING DATE: 2000-04-26
; PRIOR APPLICATION NUMBER: 60/199,880
; PRIOR FILING DATE: 2000-04-26
; PRIOR APPLICATION NUMBER: 60/218,591
; PRIOR FILING DATE: 2000-07-17
; PRIOR APPLICATION NUMBER: 60/271,814
; PRIOR FILING DATE: 2001-02-27
; PRIOR APPLICATION NUMBER: 60/215,855
; PRIOR FILING DATE: 2000-07-03
; PRIOR APPLICATION NUMBER: 09/839,446
; PRIOR FILING DATE: 2001-04-19
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
; LENGTH: 4821
; TYPE: DNA
; ORGANISM: Unknown Organism
; FEATURE:
; OTHER INFORMATION: Description of Unknown. Organism: POLYX
; OTHER INFORMATION: 297832_B.0.707

US-09-898-570-11

Query Match 28.6%; Score 1069.2; DB 10; Length 4821;
Best Local Similarity 63.1%; Pred. No. 0;
Matches 1836; Conservative 0; Mismatches 993; Indels 8; Gaps 9;

QY 196 CCGCGGCGCCAGGAGATGTAGATGAGTGTGCCCCAAGGGCTAGATGACTGCCATGCCG 255
DB 4412 CCCCTCTGCCCTGCCAGATGTGATGAGTGTGAGAGGGGACTGACAATGCCACATCG 4353
QY 256 ACGCCCTGTGTGAGACACACACCTCCTACAGTGTCTCTCAGCGCTGGCTACCAAG 315
DB 4352 ATGCTATCTGCCAGAACCCCGAGGTCTACAGTGTCTCTCAGCGCTGGCTACCAAG 4293
QY 316 GGGAGGCGAGCGAGTGTGAGGACATCCGATGAATGTGAAATGAGCTCAATGAGGCTGTG 375
DB 4292 GGGAGGCGAGCGAGTGTGAGGACATCCGATGAATGTGAAATGAGCTCAATGAGGCTGTG 4233
QY 376 TCCATGACTGTGTGAATATTTCCAGGGAATATGTTGCACTTGTCTTGTAGTGGCTTCAATG 435
DB 4232 TSCATGACTGTGTCAACATCCCTGGCAATACCGGTGTACCTGCTATGATGATGATGCCACC 4173
QY 436 TGGCTCATGACGGTCAATATTTCTGATGTGGACGAGTGCCTGGAGAAACAATGGCGGCT 495
DB 4172 TGGCACAATGACGGGACCAACTGTCTGGATGTGGACGAGTGTGCCGAGGGCAACGGCGGCT 4113
QY 496 GCCAGCATACCTGTGTCAACGCTCATCGGAGGACTATGATGCTGTGCAAGGAGGGGTTTT 555
DB 4112 GTACGAGAGCTGTGTCAACATGATGGGAGCTATGATGCTGCACTGCGGGAAGGCTTCT 4053
QY 556 TCCTGAGTGACATTCAGACACCTGCAATTCACCGCTCGGAGAGGGGCTGAGCTGCATGA 615
DB 4052 TCCTCAGCGACACCACTGATTCAGGCGGCGGAGAGGAGGAATGAAATGCAATGA 3993
QY 616 ATAGGATCAGCGCTGTGATGACATCTGCAAGGAGGCGGCAAGGGGAGGCTGCCCTGTG 675
DB 3992 ACAAGAACCAAGGCTGTGCCACATTTCCGAGGAGACACCCAGGGGGGTATTCCTCTGTG 3933
QY 676 AGTCAGGCGCTGTGTTTGTGCTGCCCAAGAACCCAGAGAGACTGCTATTTGACCTGTAAAC 735
DB 3932 AATCCGCTCTGCTGCTTGTGCTTACCAAGAACCAACCGGACTGTAAATTCACATGCAACT 3873
QY 736 ATGGGAAGCTGGGTGCCAGACTCTCTGCAACATACAGCCGATGGCCCGAGAGTGCAGCT 795
DB 3872 ATGGTAACGCGGCTGCCAGCACACGCTGTGATGACACAGAGAGGCTCCCGGCTGCCGCT 3813
QY 796 GCCATCCACAGTCAAGATGCACACAGATGGGAGGCTGCCCTGAGCGAGAGACACTG 855
DB 3812 GCCATATCAAGTTTGTGCTCCATACCGAGGGAAGACATGCAATCGGG----- 3766
QY 856 TCCTGAGGTACAGAGAGCAACACACATCAGTGTGGATGGGATAAAGCGGTGAAC 915
DB 3765 ----GAAAGCGGCTAGAGCAGCACATCCCACT-----C 3735
QY 916 GCGGCTGCTCATGGAAACGCTGTGCTGTCAACAAATGAGAGCTGTGACCGACCTGTAAAG 975
DB 3734 AAGCGCTTCTTAATGAGACCTGTGCTGTCAACAAACCGGGGCTGTGACAGTAAGTGCCATG 3675
QY 976 ATACTTCAGAGGTGTCCAGTGGAGTTGCTCTGTGGATTCACTCTCCAGTTGATCGGA 1035
DB 3674 ATGACGCACTGTGTCTCCACTGCACCTGCCCTGTGGGCTTCACTGTGACCGAGAGGGA 3615
QY 1036 AGCATGTAAAGATATTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1695
DB 3614 AGAGTGCAGAGATATAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3555
QY 1696 AAGACATCGTGGGAGCTTTTGTGCTGCGGCTGCAAGAAAGGATTTAAATTTTACAGATG 1155
DB 3554 GCAACACAGTGGGAGCTTCGAAATGCAAGTTGCAAGAAAGGCTTAAAGCTTCTCATCATG 3495
QY 1156 AGAGTCTTCCAGAGATGTGGATGAGTGTCTTGTGATAGAGCTGTGACACAGCTGCA 1215
DB 3494 AGAGAACTGCCAGAGATATAGACGAGTGTCTTGTATGAACTGTGACACATATGTG 3435
QY 1216 TCAACACCTGCGACATTTTGTGCTGCTGCAACCGAGGATACACCTGTATGGCTTCA 1275
DB 3434 TCAACACCTGCGAGAGCTTCCAGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3375
QY 1276 CCCATGTGGAGACACCAATGAGTGCAGCATCAACAAACGGAGGCTGTGACGAGGCTCTGTG 1335

3374 CCCACTGTGGGATGTGGATGAATGAGCATCAACGGGGAGGTTGCCGCTTTGGCTGCA 3315
1336 TGAACACAGTGGGAGCTATGATGCGAGTGCACCCCTGG---GTACAGGCTCCACTGGA 1392
3314 TCAACACTCCTGGCAGCTACCAAGTGTACTGTGCCAGCAGCGCAGGGTGGCTGSCACTGGA 3255
1393 ATAAAAAGACTGTGTGGAAAGTGAAGGGGCTCTCTGCCACAAAGTGTGTCAACCCCTGTGT 1452
3254 ATGGCAAGATTGCACAGAGCACTGAAGTGTACGGGCACTCTGGGCTCTGGAAGCCA 3195
1453 CCTGCACTGCGGTGAAGTGTGTGAGAGAGCGGTGCTTCTCAGATGTCACTCTGCA 1512
3194 TCGTCACTGCAACCGGTCTGGCAAGAGGACACCTGTGTCCTGACCTGTCTCCACGGG 3135
1513 TTCACCTCTCTCAGA-----TGTCAACCACTCAGGACAGTGTACCTTTAGCTAA 1566
3134 CCGGATTTTGCAGAGGCTGCAGTGTCTCATTAAACACGGGCTCTCTTCAAGATCA 3075
1567 ATGAAGGCAAGTGTAGTTTGAA-----AAATGCTGAGCTGTTTCCCGAGGGTCTGCCAC 1620
3074 AGGATGCCAAATGCCGTGTGCACCTCGGAACAAGGGCAACAGAGGAGGTGGCAGAA 3015
1621 CAGCACTACCAAGAGACACAGCTCAGTAAAGAGAGCTTCGCTACGTAACCTTAGAT 1680
3014 CCACAGGGCCAGGTGTGTGCCCTCTCTGAATGCCAGGTCACTTCACTCACTTAAGT 2955
1681 CGAGCTCTGGCAAGCAAGTCCCGAGAGCCCTGGCGGACCAAGACCCCTTAGGAATGT 1740
2954 GTGACTTCTCTGGAAAGGGCAAGGGCCGACGGGCCGGACCCCTCAGGCAAGAGGTCA 2895
1741 TTA---TCACTGTGAGTTTGAGCTTTGAAACTAAACAAAGAGAGTGCAGAGCTTTTGTG 1797
2894 CAAGGCTCACTCGGAACCTGGAGGAGAGGTTCAGAGCCGGAAGAACACAGCCAGCTGTG 2835
1798 ACTGAGTGCATCTGTAAGAGCAACGAGAGAGGCTCGTAAAGCATCCGACAGCTCA 1857
2834 GGCTGCCCTGCTCCACAGCGAAATGGAACGGCGGCTGAAAGGATCCCTGAAGATGCTCA 2775
1858 GAAAGCCGTCCACAGGAGAGCTTTCACCTCCAGCTCTCAGGATGAACCTCGACGTGG 1917
2774 GAAAGTCCATCAACAGGACCGCTTCTGCTGGCTGGCAGGCTTGATATGAGCTGG 2715
1918 CTAAAAAGCCTCCCAAGACATCTGAACGCCAGGCA-----GAGTCTCTGTGAGTGG 1968
2714 CCCACAAGCGCGCTGTGTAGCGGGGAGCGAGAGCGGATGAGTCTGTAGCGCGC 2655
1969 GCCAGGTCTACAGAAACCAATGTGTGAGTGTGAGGCTGGGACCTATTATGATGGAG 2028
2654 GGCAGCACCGTGTGGGACCAAGTGTGTGAGTGTGCGCGGAGGAAAGCTATTACCAAGGCC 2595
2029 CACGAGAACGCTGCATTTTATGTCCAAATGGAACCTTCCAAATGAGGAGGACAAATGA 2088
2594 AGACGAGAGTGTGTGCCATGCCAGCGGCACTTCCAGAGAGAGAGAGGCGAGCTCT 2535
2089 CTGTGAAACCATGCCCAAGACCAAGAAATCTGGGSCCTTGAGACCCCAAGCTTGA 2148
2534 CCTGCACTCTTGCCTGGGA-----GTGATGCCCAAGCGGCTCTTGGAGCCACA 2484
2149 ATATGTCTGAATGTGAGAGTGTGTCAACCTGGTGAATATTCTGCAGATGCTTGCAC 2208
2483 AGCTACCAAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 2424
2209 CTGCCAGCTGTGTGCCCTGGGACGCTTCCAGCCTGAAGCTGGTGAACCTCTCTGCTTCC 2268
2423 CTTGTGAGCATGCCACAGCTGGCACTTACACCTGAAGCAGGAGGACCCCTATGCTTCC 2364
2269 CTTGTGAGGAGGCTTGGCAACCAACATCAGGAGACTACTTCTCTTTCAGGAGCTGAAA 2328
2363 CTTGTGGTGGGSCCTCACCACCAAGCATGAAGGGGCCATTTCCTTCAAGACTGTGACA 2304
2329 CAGAGTTCAATGTTCACTGAGCATTTCTACAAACACCACTCAGCATGATTTCTGTT 2389

2303 CCAAAGTCCAGTCTCCCCAGGCGCACTACTTCAACACCAAGCATCCACCGCTGTATTGGCT 2244
2389 GCCCAGTGGGAAACATACAGCCTGAATTTGGGAAAAAATAATTTGTGTTTCTTGCCCAAGAA 2448
2243 GTGGCAATGGGCTCTCTACAGCCGACTTTCGCTGAGAACTTCTGACGCGCTGTCCAGGAA 2194
2449 ATACTAGACTGACTTTGATGGCTCCCAAAACATAACCCAGTGTAAAAAAGAGAGATGTG 2509
2183 ACACAGCACAGACTTTGATGGCTTACAGTGTGGCCCAATGCAAGATCTGTCAGTGTG 2124
2509 GAGGGAGCTGAGAGATTTCACTGGGTACATTGAAATCCCAAACTACCCAGGCAATTTACC 2568
2123 GTGGGAGCTGGTGAATTTCACTGGGTATTTAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 2064
2569 CAGCAACACAGGAGTGTACGTGGAGCATGACCCAGCCAGCCCAAGCGCGCATCTGATCG 2628
2563 CAGTGTGTGGAGTGTGATCTGGAACATCAACCCCTCACCACAGCGCAAGATCTTATCG 2004
2529 TGGTCCCTGAGATCTTCTCCCTCCATAGAGAGCACTGTGGGAGCTATCTGCTGATCGGA 2688
2503 TGGTACCGAGAGATCTTCTGCCATCTGAGGATGAGTGTGGGAGCTCTCTGTCATGAGAA 1944
2689 AAACCTCTTCACTGATTTCTGTGACAAACATATGAACCTGCCAGACCTAGCAAGCGCCCA 2748
1943 AGAATCATCCCATCCCTCCATTTACCACTTATGAGACCTGCCAGACCTACGAGCGTCCA 1894
2749 TCGCTTCACTCCAGTCAAGAGCTGTGGATTGAGTTCAAGTTCAGTTCAATGAAGGGAACA 2808
1893 TTGCTTCACTGCGCTTCCAGGAGCTTCCAGGAGCTTGGATCACTTCAAGCAAGCGGCGCA 1824
2809 GCGTAGAGGTTTCCAGTCCCATAGTGCATATGATGAGAGCTACCAAGAACTCATTTG 2868
1823 GCGCCCTGGCTTCCAGATTTCCCTATGTTACCTATGATGAGGACTATGAGCAGCTGTAG 1764
2869 AGACATAGTTCAGATGGCAGGCTCTATGCTGTCAGACCATCAGGAAATCTTTAAGG 2528
1763 AAGACATGTGCGAGATGGCGGCTCTATGCTCTGAAATCCACCAAGGAGATTTAAAGG 1704
2529 ATAGAACTTATCAAGGCTCTCTTTGATGCTCTGGCCCATCCCAAGAACTATTTCAAGT 2988
1703 ACAGAGAGCTCATGAGGCTTCTTTGAGTGTGAGGCTAGCCACCCCAAGAACTACTTCAAGT 1644
2389 ACACAGCCAGGAGTCCGAGAGATGTTTCCAGAGATGTTTCCAGAGTGTTCATCCGATTCAGTTCCA 3048
1643 ACACAG---AGAAACAAGAGGATGCTGCAAAATCTTTCATCAAGCTGCTCCGCTCCA 1587
3049 AGGTGTCAGGTTTTCAGACCTTACAAT 3078
1586 AAGTTCCAGCTTCTGAGGCGCTTACAAAT 1557

RESULT 7

US-09-839-446-11/C
; Sequence 11, Application US/29839446
; Publication No. US2003050232A1
; GENERAL INFORMATION:
; APPLICANT: GERLACH, VALERIE L.
; APPLICANT: ELLERMAN, KAREN
; APPLICANT: MACDUGALL, JOHN R.
; APPLICANT: SMITHSON, GLENDA
; TITLE OF INVENTION: NOVEL HUMAN PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND
; TITLE OF INVENTION: METHODS OF USING THE SAME
; FILE REFERENCE: 15966-776
; CURRENT APPLICATION NUMBER: US/09/839,446
; CURRENT FILING DATE: 2001-04-19
; PRIOR APPLICATION NUMBER: 60/198,293
; PRIOR FILING DATE: 2000-04-19
; PRIOR APPLICATION NUMBER: 60/198,645
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: 60/210,809
; PRIOR FILING DATE: 2000-06-09
; PRIOR APPLICATION NUMBER: 60/199,476
; PRIOR FILING DATE: 2000-04-26

```
; PRIOR APPLICATION NUMBER: 60/200,025
; PRIOR FILING DATE: 2000-04-26
; PRIOR APPLICATION NUMBER: 60/224,610
; PRIOR FILING DATE: 2000-08-11
; PRIOR APPLICATION NUMBER: 60/200,024
; PRIOR FILING DATE: 2000-04-26
; PRIOR APPLICATION NUMBER: 60/199,880
; PRIOR FILING DATE: 2000-04-26
; PRIOR APPLICATION NUMBER: 60/218,591
; PRIOR FILING DATE: 2000-07-17
; PRIOR APPLICATION NUMBER: 60/271,814
; PRIOR FILING DATE: 2001-02-27
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: Patencin Ver. 2.1
; SEQ ID NO 11
; LENGTH: 4821
; TYPE: DNA
; ORGANISM: Unknown Organism
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: POLYX
; OTHER INFORMATION: Z97832.B.0.707
US-09-839-446-11

Query Match      28.6%; Score 1069.2; DB 11; Length 4821;
Best Local Similarity 63.2%; Pred. No. 0;
Matches 1836; Conservative 0; Mismatches 993; Indels 81; Gaps 9;

QY 196 CCGCGGGGCGGAGGAGGATGATGAGTGTGCCCAAGGGGTAGATGACTGCCATGCCG 255
DB 4412 CCCCTCTGCCCTGCCAGATGGATGAGTGTGGAGGGGACTGACAACTGCCATCG 4353
QY 256 AGCCCTGTGTGAGAACACACCCACCTCTCTAAGTGTCTTGCAGGCTGCTGCCAAG 315
DB 4352 ATGCTATCTGCCAGAACACCCGAGGTCATACAAGTGCATCTGCAAGTCTGGCTACAG 4293
QY 316 GGGAGGAGGAGGAGTGTGAGGACATCGATGAATGTGGAAATGAGCTCAATGAGGCTGTG 375
DB 4292 GGGAGGAGGAGGAGTGTGAGGACATCGATGAATGTGGAAATGAGCTCAATGAGGCTGTG 4233
QY 376 TCCATGACTGTTTGAATATTCAGGCAATTTATCGTTGCACTTTTGTATGGCTTCAATGT 435
DB 4232 TGCATGACTGTGTCAACATCTCCCTGGCAATTAACGGTGTACCTGTATGATGATTCACCC 4273
QY 436 TGGCTATGACGGCTATATTTGTCTTGAATGTGACACGAGTGCCTGAGAACAAATGGCGCT 495
DB 4172 TGGCAGATGACGACACAACTGTCTGSAATGTGACAGTGTGCCAGGGCAACGGCGCT 4113
QY 496 GCGAGCATACCTGTGTCAACGTCTATGGGAGCTATGAGTGTCTGCAAGGAGGGGTTT 555
DB 4112 GTCAGAGAGCTGTGTCAACATGATGGGAGCTATGAGTGTCTGCAAGGAGGGGTTT 4053
QY 556 TCCATGAGTGAACATACGACACCTGCAATTCACCGCTCGGAAGAGGGCTGAGCTGATGA 615
DB 4052 TCCATGAGTGAACATACGACACCTGCAATTCACCGCTCGGAAGAGGGGTTT 3993
QY 616 ATAAGATCAGGCTGTATGATCATCTGCAAGAGAGGCCCCAAGGGGCAACGCTGCTGTG 675
DB 3992 ACAAGAACACGCTGTGCCACATTTTGGCGGAGACACCCCAAGGGGGGTTTGTGCTGTG 3933
QY 676 AGTGAGGCTGTGTTTGTAGCTGGCCAGAACACAGAGAGCTGATCTTGTAGCTGTACC 735
DB 3932 ATGCGCTGTGCTGTGAGCTTACCAAGAACCAACGGGAGCTGTAATTTGATCACTCACT 3873
QY 736 ATGGGAACGCTGGTGGTGGCAGCACTCTCTGTGACATACAGCCGATGCGCCAGAGTGCAGT 795
DB 3872 ATGGTAACGCGCTGCCAGCACAGCTGTGATGACACAGAGAGGGTCCCCGCTGCGCT 3813
QY 796 GCATCCACAGTACAGATGACACAGATGGAGGAGCTGCTTGTAGGAGAGGACATG 855
DB 3812 GCCATATCAAGTTTGTGCTCCATACCGAGGGAAGACATGCTCGGG----- 3766
QY 856 TCTGTGAGGTGACAGAGAGCAACACCATCATGTTGTGTGATGGGATTAACGGGTGAAC 915
```

```
DB 3765 -----GAAAGGGGCTAGAGCAGACATCCCACT-----C 3735
QY 916 GGGCGCTGCTCATGGAAACGTGTCTGTCAACATGAGGCTGTGACCGCACTGTGAAG 975
DB 3734 AAGCGCTTCTATAGAGACCTGTCTGTCAACACGCGGCTGTGACAGTAAGTGCATG 3675
QY 976 ATACTTGACAGAGTGTCTCACTGAGTGTCTGTGATGATCACTCTCCATTTGATGGGA 1035
DB 3674 ATGAGGAGCTGTGTCTCACTGCACTGCACTGCGCTTCACTGCGAGGAGGAGG 3615
QY 1036 AGACATGTAAGATATTTGATGAGTGCACAGCCGCAATGAGGTTGTGATCATTTCTGCA 1095
DB 3614 AGAGTGCATAAGATATAGATGAGTGCCTTAAACACGGGGCTGTGACATATTTGCC 3555
QY 1096 AAAACATCTGGGCGAGTTTACTGCGGCTGCAAGAAAGGATTTAAATATTAAACAGATG 1155
DB 3554 GCACACAGTGGGCGAGCTTCCGATGAGTTGCAAGAAAGGCTATAAGCTTCTCATCAATG 3495
QY 1156 AGAAGCTTTCACAGATGTGATGAGTGTCTTGGATAGAGCTGTGACACAGCTGCA 1215
DB 3494 AGAGAACTGCCAGGATATAGACGAGTGTCTTGTGATCGAACTGTGACACATATGTG 3435
QY 1216 TCACACACCTGGGCAATTTGCTGTGCTTGCACACCGAGGCTACACCTGTATGGCTCA 1275
DB 3434 TCACACACAGGAGGCTTCCAGTGTCTTGCATGCTGTGCTGTGATGATGATCA 3375
QY 1276 CCCACTGTGGAGACACCAATGAGTGCAGATCAACACGGAGGCTGTGACAGTCTGTG 1335
DB 3374 CCCACTGTGGGATGTGATGATGAGTGCAGATCAACCGGAGGTTTGGCTTGTGCTGCA 3315
QY 1336 TGAACAGTGGGAGGCTATGATGAGTGCAGTGCACCTGTG---GTACAAGTCCACTGGA 1392
DB 3314 TCAACACTCTGGCAGCTACAGTGTACCTGCGCAGAGGCAAGGCTGCGCTGCACTGGA 3255
QY 1393 ATAAAAAGACTGTGTGAAGTGAAGGGCTCTGTGCCCAACAGTGTGTCAACCGCTGTG 1452
DB 3254 ATGCAAGATTCACAGAGGCTGAGTGTGAGGCTGTGAGGCTGTGAGGCTGTGAGGCTG 3195
QY 1453 CCCTGCACTCGGTAAAGTGTGAGGAGAGGGGCTTCTCTCAGATGTCACTCTGGCA 1512
DB 3194 TGCTCAGCTGCAACCGGTGTGGCAAGAGGACACCTGTGCGCTGACCTGTCCCTCCAGG 3135
QY 1513 TGCACCTCTCTGAGA-----TGTACACACCATCAGACCAAGTGTAACTTTAGCTAA 1566
DB 3134 CCCGATTTGCCAGAGGCTGCACTGTGCTTCAATTAACCAACGGGCTCTCTCAAGATCA 3075
QY 1567 ATGAAGCAAGTGTAGTTTGA-----AAATGCTGAGCTGTTTCCGAGGCTGTGCGAC 1620
DB 3074 AGGATGCCAATGCGTTTGCACCTGCGAAACCAAGGCAACAGAGGAGGCTGGCGAA 3015
QY 1621 CAGCACTACAGAGAGGACAGCTCAGTAAAGAGAGCTTCCGCTAGCTAAACCTTACAT 1680
DB 3014 CCACAGGCGCAGTGTGCGCTGCTGTGAATGCCAGGTCACTTCACTTCACTTAACT 2955
QY 1681 GCASCTGTGGAGAGCACTGCGAGGAGGCTGCGCAGCAGCAAGCAGCCCTTAAGAAATGT 1740
DB 2854 GTSACTCTCTCGAAGGCGAAGGCGCGAGCGGCGCGACCCCTCCAGGCAAGAGGCTCA 2895
QY 1741 TTA---TCACCTGTGAGTTTGAAGTGTGAACATTAACAAAAGGAGGCTGAGCTTCTGTG 1797
DB 2894 CAAGGCTACCTCGAAGTGTGAGGCGAGGCTCAGAGCGGAGAAACACAGCAGCTGTG 2835
QY 1798 ACCTGAGCTGCATCGTTAAGCGAAACCGAGAGCGGCTTCCGTAAAGCCATCCGACGCTCA 1857
DB 2834 GGTGCTGCTGCTCCGACAGCAAGTGTGAACGGGCTGAAAGGATCTCTGAAGATGCTCA 2775
QY 1858 GAAAGGCGCTCCACAGGAGGAGGTTTCACTTCCAGCTCTCAGGCAATCAACCTCAGCTGG 1917
DB 2774 GAAAGTGCATCAACAGGAGGCTTCTGCTGCTGCGCTGCGAGGCTTGAATATGAGCTGG 2715
QY 1918 CTAAAGGCTCCCGAGAACATCTGAACCGCAGGCA-----GAGTCTGTGGAGTGG 1968
DB 2714 CCCAAGCGCGGCTGTGTAGCCGGAGGAGCAGAGCGATGAGTCTGTAGGCGCG 2655
```

QY	1969	GCCAGGGTCATGCAGAAAA	CAATGTGTCTCAGTTGCAGGGCTGGGAACCTATTATGATGGAG	2028
DB	2654	GGCAGCACCGTGTCTGGGACCAAGTGTCTCAGCTGCCCGCAGGGAACGTATTATCCACGGCC	2595	
QY	2029	CACGAGAACGCTGCATTTTATGTCCAAATGGACCTTC	2089	
DB	2594	AGACGGAGCAGTGTGTCCATGCCCCAGCGGCACTTCCAGGAGAGAAAGGCAAGCTCT	2535	
QY	2089	CTTGTGAACCATGCCCAAGACCCAGGAAATCTGGGGCCCTG	2148	
DB	2534	CTTGGACCTTTGCCCTGGGA-----GTGATGCCACGGGCGCTCTTTGGAGCCACCA	2484	
QY	2149	ATATGCTGAATGTGGAGGCTGTGTCAAACTGGTGAAATTTCTGCAGATGGCTTTTSCAC	2208	
DB	2483	ACGTCAACCGTGTGCAGGTGAGTGCACCTGGCCCAACACTCTGTAGAAGGTTCAAGC	2424	
QY	2209	CTTGCCACGCTCTGTGCCCTGGSCAGTTTCAGGCTGAAGCTGGTGGMACTTCTGCTCTCC	2269	
DB	2423	CTGTTCAGCCATGCCCAAGTGGCACTTACCAACTGAACGAGGACGACCTATGTGTTCC	2364	
QY	2269	CTGTGGAGGAGGCTTGGCCACCAACATCAGGGAGCTACTTCTTTTCAGGACTGTGAAA	2328	
DB	2363	CTTGTGTGGGGCTCACCAACAGCATGAAGGGGCAATTTCTTTCCAGACGTGTGACA	2304	
QY	2329	CCAGAGTTCAATGTTTCACTCTGGACATTTTCACAAACACCACTCATCCGATGTAATTCGT	2385	
DB	2303	CCAAAGTCCAGTGTCTCCAGGGCACTACTACAAACACGACATCCACGGCTGTATTCGCT	2244	
QY	2389	GCCAGTGGGAAATACACAGGCTGAATTTGGAAAAATAATTTGTGTTTCTTGCCAGGAA	2448	
DB	2243	GTGCATGGGCTCCTATCAGCCGCACTTCGGTCAGAACTTTCTGCAGCGCGTGTCCAGGAA	2184	
QY	2449	ATACTACGACTGACTTTGATGGCTCCACAAACATAAGCCAGTGTAAACACAGAAGATGTG	2508	
DB	2183	ACACAACACAGACTTTTGTATGGCTTCTACAGTGTGGCCCAATGCMAAGATCGTCACTGTG	2124	
QY	2509	GAGGGAGCTGGGAGATTTTCACTGGGTACATTTGAATCCCCAACTACCCAGGCAATPACC	2568	
DB	2123	GTGGGAGCTGGGTGAGTTCACCTGGCTATATTGAGTCCCCCACTACCCGGGCAACTACC	2064	
QY	2569	CAGCCAAACCGAGTGTACGTGGACCATCAACCCACCCCGACGGCGGATCTCTGATCG	2628	
DB	2063	CAGCTGTGTGGAGTGCATCTGGAAACATCAACCCCCCACCCAGCGCAAGATCTTATATCG	2004	
QY	2629	TGGTCCCTGAGATCTTCTCGCCATCAGAGACGACTGTGGGACTATCTGGTGATGGGA	2688	
DB	2003	TGGTACCAGAGATCTTCTCGCATCTCAGGATGAGTGTGGGAGCTCTCTGTCTATGAGAA	1944	
QY	2689	AAACCTCTTCATCCAAATTTGTGACACATATGAAACCTGCCAGACCTACGAAAGCCCA	2748	
DB	1943	AGAACTCATCCCCATCTCCATTTACCATTATGAGACCTGCCAGACCTACGAGCGTCCCA	1884	
QY	2749	TCGCCTTTCACCTCCAGGTCAAGAGAGCTGTGGATTCAGTTCAAGTCCAATGAAGGGAA	2808	
DB	1883	TTGCTTTCACCTGCCCGTTCAGGAAAGCTCTGGATCAACTCAAGACAAGCGAGGCCAAC	1824	
QY	2809	CGCGTAGAGGGTTCCAGGTCCCATCGTACATATGATGAGGACTACGAGAACTCAATG	2868	
DB	1823	CGCGCGTGGCTTCCAGATTTCCATATGTACCTATGATGAGGACTATGAGCAGCTGTGAG	1764	
QY	2869	AAGACATAGTTTCAGATGGCAGGCTCTATGCATCTGAGAACCATCAGGAAATATCTTAAAG	2928	
DB	1763	AAGACATTTGCGAGATGGCGGCTCTATGCCTCTGAAACACCAACGAGGAGATTTTAAAGG	1704	
QY	2929	ATAAGAAACTTATCAAGGCTCTGTTGATGTCTTGGCCCATCCCGAGAACTATTTCAAGT	2988	
DB	1703	ACAAGAAGCTCATCAAGGCTCTTCTTGGAGTGTCTAGGCCACCCCGAGAACTTCTCAAGT	1644	
QY	2989	ACACAGCCAGGAGTCCCGAGAGATGTTTCCAGATGCTTCATCCGATTCGTACGTTTCCA	3048	
DB	1643	ACACAG---AGAAACACAAGGAGATGCTGCCAAATCTTTCATCAAGGTGCTGCGTCCA	1587	

```

01 3049 AAGTGTCCAGCTTTTTCAGACCTTACAAAT 3278
02 ||||| ||||| ||||| ||||| |||||
03 1586 AAGTTTCCAGCTTCTGAGCGCCTACAAAT 1557
04 ||||| ||||| ||||| ||||| |||||

RESULT 8
US-89-898-570-13
: Sequence 13, Application US/09898570
: Patent No. US2002C123612A1
: GENERAL INFORMATION:
: APPLICANT: GERLACH, VALERIE L.
: APPLICANT: ELDERMAN, KAREN R.
: APPLICANT: MADDOUGALL, JOHN R.
: APPLICANT: SMITHSON, GLENDA
: TITLE OF INVENTION: NOVEL HUMAN POLYHUMAN PROTEINS, POLYHUMAN
: TITLE OF INVENTION: METHODS OF USING THE SAME
: FILE REFERENCE: 15966-776CIP
: CURRENT APPLICATION NUMBER: US/89/898,570
: CURRENT FILING DATE: 2001-07-03
: PRIOR APPLICATION NUMBER: 60/198,293
: PRIOR FILING DATE: 2003-04-19
: PRIOR APPLICATION NUMBER: 60/198,645
: PRIOR FILING DATE: 2003-04-20
: PRIOR APPLICATION NUMBER: 60/210,809
: PRIOR FILING DATE: 2003-08-29
: PRIOR APPLICATION NUMBER: 60/199,476
: PRIOR FILING DATE: 2003-04-26
: PRIOR APPLICATION NUMBER: 60/200,025
: PRIOR FILING DATE: 2003-04-26
: PRIOR APPLICATION NUMBER: 60/224,610
: PRIOR FILING DATE: 2003-08-11
: PRIOR APPLICATION NUMBER: 60/200,024
: PRIOR FILING DATE: 2003-04-26
: PRIOR APPLICATION NUMBER: 60/199,880
: PRIOR FILING DATE: 2003-04-26
: PRIOR APPLICATION NUMBER: 60/218,591
: PRIOR FILING DATE: 2003-07-17
: PRIOR APPLICATION NUMBER: 60/272,914
: PRIOR FILING DATE: 2001-03-27
: PRIOR APPLICATION NUMBER: 60/215,955
: PRIOR FILING DATE: 2003-07-03
: PRIOR APPLICATION NUMBER: 09/839,446
: PRIOR FILING DATE: 2001-04-19
: NUMBER OF SEQ ID NOS: 59
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 13
: LENGTH: 4550
: TYPE: DNA
: ORGANISM: Unknown; Organism
: FEATURE:
: OTHER INFORMATION: Description of Unknown Org:
: OTHER INFORMATION: Z97832_B1
: US-89-898-570-13

```

Qy	416	TTGTTTGTATGGCTTCATGTTTGGCTCATGACGGTCATAATTGCTTCTGATGTGACAGAGTG	475
Db	359	CTGCTATGATGGATTTCCACTGGCACAATGACGGACACAACATGTCTGATGTGGACAGTG	418
Qy	476	CTTGGAGAACAAATGGCGGCTCCAGACATCCTGTGTCAACGTCTATGGGAGCTTATGAGTG	535
Db	419	TGCCGAGGGCAACGGCGGCTGTGACAGAGCTGTGTCAACATGATGGCAGCTATGAGTG	478
Qy	536	CTGCTGCAAGGAGGGGTTTTTCTGTAGTGACAAATCAGCACACCTGCAATTCACCGCTCGGA	595
Db	479	CCACTGCCGGGAAGGCTTCTCTCAGCGACACACAGACATACCTGTATCCAGCGGCCAGA	538
Qy	596	AGAGGGCTGAGCTGCATGAATTAAGGATCAGGCTGTAGTCACATCTGCAAGGAGAGCCCC	655
Db	539	AGAAGGAAATGAATTTGCATGTAACAAAGAACCAACGAGCTGTGCCCAATTTGCCGCGAGACAC	598
Qy	656	AAGGGGACAGCGTCCGCTGTGAGTCGAGGCTGTGTTTGTAGCTGCGCAAGAACACAGAGAGA	715
Db	599	CNAGGGGGGTATTTGCCGTGTGAATGCCGCTCTGTGCTTTGAGCTTACCNAGAACCAACGGGA	658
Qy	716	CTGCATCTTCACTGTAAACCATGGGAACGGTGGGTGCCAGCATCTCTGTGACGATACAGC	775
Db	659	CTGTAAATTCACATGCAACTATGCTAACGGCGGCTGCCAGCACACGCTGTATCACACAGA	718
Qy	776	CGATGGCCACAGCTGCAGCTGCCATCCACAGTACAGATGCACACAGATGGGAGGAGCTG	835
Db	719	GCAGGTTCCCGGTGCGGCTGCCATATCAAGTTTGTCTCCATACCGACGCGGAAGACATG	778
Qy	836	CTTTGACGAGAGGACACATGCTCTCTGGAGGTGCACAGAGACCAACACACATCAGTGTGTGA	895
Db	779	CATCGGG-----GNAAGGCGCTAGACGAGCACATCCCCACT-----	815
Qy	896	TGGGGATAACGGGTGAAACGGCGGCTGCTCATGGAACGTGTGCTGTCAACAAATGGAGG	955
Db	816	-----CAAGCCGTTTCTAATGAGACCTGTGCTGTCAACAAACGGGG	856
Qy	956	CTGTGACCGCACCTGTAAGGATACCTCGACAGGTGCCACTGCAGTTGTCTCTGTGGATT	1015
Db	857	CTGTGACAGTAAGTGCCATGATGCGGAGCTGGTGTCCATGCACCTGCCCTGTGGCTT	916
Qy	1016	CACCTCTCCAGTTGGATGGGAGACATGTAAGATATTTGATGAGTGGCCAGACCGCGCAATGG	1075
Db	917	CATGCTGCACCCAGACAGGAAGAGCTGCAAGATATAGATGAGTGGCGCTTAACACACGG	976
Qy	1076	AGTTGTGATCATTTTGTGCAAAAACATCGTGGGAGCTTTTGTACTGCGGGCTGCAAGAAAGG	1135
Db	977	GGGCTGTGACCATATTTGCCGCAACACAGTGGGAGCTTCSAATGCAAGTTGCAAGAAAGG	1036
Qy	1136	ATTAAATATTACACATGAGAAGCTTCCCAAGATGTGATGAGTGTGCTTTGTGATAG	1195
Db	1037	CTATAAGCTTCTCATTAATGAGGAAGTGCACAGATATAGAAGATGTTTCTTTTGTATCG	1096
Qy	1196	GACCTGTGACCCAGCTGCATCAACACCCCTGGCACATTTTCTTGTGCTTCAACCCAGG	1255
Db	1097	AACCTGTGACCAATAATGTGTCAACACACAGGAAGCTTCCAGTGTCTCTGCCATCGTGG	1156
Qy	1256	GTACACCTGTATGCTTCAACCACTGTGAGACACCAATGAGTGCAGCATCAACACAGG	1315
Db	1157	CTACCTGTTGTATGTTATCAACCACTGTGGGATGTGGATGAATGACAGATCAACCGGG	1216
Qy	1316	AGGCTGTGACGAGCTGTGTGAAACAGATGGGAGCTATGAATGCCAGTGGCCACCTGG	1375
Db	1217	AGGTTGCGGCTTTGGCTTGATCAACATCTCTGCGAGCTACCAAGTGTACCTGCCAGCAGG	1276
Qy	1376	---GTAACAAGCTCCATGGAAATAAAAAGACTGTGTGGAAGTGAAGGGGCTCTCTGCCAC	1432
Db	1277	CCAGGGTCGGCTGCACTGGAAATGGCAAGATTTGCACAGAGCCACTGAAGTGTGAGGCGAG	1336
Qy	1433	AAGTGTGTCAACCGCTGTGTCCCTGCACTGGCGTAAAGTGTGGTGGAGGACAGGGTGCTT	1492
Db	1337	TCCTGGGGCTCGAAAGCCATGTCTCAAGTGTCACTGCACCTGCAACCGCTCTGGCAAGAGGACACTGTGC	1396

Qy	1493	CCTCAGNTGTCACTCTGGGCAATTCACCTCTCTTCAGATGTGACCA-----CCATCAGGAC	1544
Db	1397	CCGTACCTGTCCCTCCAGGCGCGGATTTTTCGACAGGCTGTAGTGTCTGTCCATTAACA	1456
Qy	1547	AAAGTGAACCTTTAAGCTAAATGAAGGCAAGTGTAGTTTGA-----AAATCTGTAGCT	1600
Db	1457	ACGGGCTCTTTCAGATCAGGATGCCAAATGCCGTTTGCACCTCGGAACAAGAGGCA	1516
Qy	1601	GTTTCCCGAGGGTCTTCGACCCAGCACTACCAAGAAACACAGCTCAGTAAACAGAGCTT	1660
Db	1517	AACAGAGGAGGCTGGCAGAACACACGGGCCAGTGGTGGCCCTGTCTGTGAATGCCAGGT	1576
Qy	1661	CCGCTACGTAAACCTTACATCTGCAAGCTCTGCGAAGCAAGTCCGAGGAGCCCTTGGCGAGC	1720
Db	1577	CACCTTCATCCAGCTTAAGTGTGACTCTCTCGAAGGGCAAGGGCGCAGGGGCGCGGAC	1636
Qy	1721	AAGCACCCCTAAGCAATGTTTA---TCACTGTGAGTTTGAGCTTGAACATTAACCAAA	1777
Db	1637	CCCTCAGGCNAAGAGGTACAAAGCTTCAACCTTGGAACCTGGAGGAGAGGTCAAGCCGA	1696
Qy	1778	GGAGGTGCACAGCTCTTGTGACCTGAGCTGCATCTGTAAAGCAACCGAGAAGCGGCTCCG	1837
Db	1697	AGAAACACACAGCAGCTGTGGGTGCGCTCGCTCCGACAGCAATGGAACCGCGGTGAA	1756
Qy	1838	TAAAGCCATCCGACAGCTCAGAAAGCCGTCCACAGGGAGCAGTTTCACTCTCAAGTCTTC	1897
Db	1757	AGGATCCCTGAAGATGCTCAGAAAGTTCATCAACACGAGGACCGCTTCTCTCGCGCTGGC	1816
Qy	1898	AGGCATGMACTCGACGTGCTAAAAGGCTCCACGAACATCTGAACGCGCAGGCA----	1952
Db	1817	AGGCTTGAATTATGAGCTGCCCCACAAAGCCGGGCTGCTAGCCGGGAGCGAGCAGAGCC	1876
Qy	1953	----GAGTCTCTGTGAGTGGCCAGGAGTCAATGAGAAAAACCAATGTGTCAATTCGAGGCG	2008
Db	1877	GATGGAGTCTGTAGCCCGGGCAGCACCGTGTCTGGACCAAGTGTGTCACTGCCCCGA	1936
Qy	2009	TGCGACCTATTATGTAGAGCAGCAGAAAGCTGCATTTTATGTCCAAATGGAACTTCCA	2068
Db	1937	GGAAACGATTAACACGGCCAGCAGGAGCAGTGTGTGCCATGCCACCGCGGCACTTCCA	1996
Qy	2069	AAATGAGGAGGACAATGACTGTGTAACTCATGSCCTCAGACCAGGAAA--TCTGGGGCCCT	2128
Db	1997	GGAGAGAGAAGGCAGCTGTCTTCCGTGGACCTTTTSCCTCTGGAG-----GATGCCCA	2047
Qy	2129	GAAGACCCCAAGAGCTTGAATATGTCTGAATGTGGAGTCTGTGTCAACCTGGTGAATA	2188
Db	2048	CGGSCCTCTTGGAGCACCACAACTGACACCGTGTGCAGGTCACTGGCCACCTGGCCACA	2107
Qy	2189	TTCTGCAGATGGCTTTTGCACCTTGCAGCTCTGTGCCCTGGGACGTTCCAGGCTGAAGC	2248
Db	2106	CTCTGTAGATGGGTTCAAAGCCCTGTGAGCCATGCCACGTTGCCACCTGCAACTGAAGC	2167
Qy	2249	TGTTGCACTTCTCTTCTCCCTGTGAGAGGSCCTTGCCACCAACCATCAGGGAGCTAC	2308
Db	2169	AGGAGGACCCCTATGCTTCTCTTGTGTGGGGCCCTCACCAACCAAGCA--GAAGGGGCCAT	2227
Qy	2309	TTGCTTCAAGCATGTGAAACAGAGTTCAATGTTCACTTGGACATTTCTTCAACACCCAC	2368
Db	2229	TTCTTCCACAGCTGTACACCAAAATCCAGTCTCCCGAGGCCTACTATCAACACCCAG	2287
Qy	2369	CATCTACCGATGTATCTGTTTSCCATGTGGGAACATACCAAGCTTGAATTTGGAAAAATTA	2428
Db	2288	CATCCACCGCTGTATTCGTGTGCCATGGGCTCTATCAGCCGACCTTCCGTGAGAACTT	2347
Qy	2429	TTGTGTTTCTTGCCACGGAATACTACGACTGACTTTGATGGCTCCACAAAATATACCCA	2488
Db	2348	CTCAGCCGCTGTCCAGGAAACACAGCAGACACTTTTGTGCTCTTACCAAGTGTGCCCA	2407
Qy	2489	GTGTAAAAACAGAAGATGTGAGGGGAGCTGGAGATTTCACTGGGTCAATTGAATCTCCC	2548
Db	2408	ATGCAGAAATCGTCACTGTGTGGGAGCTGGGTGAGTTCATCTTGCTTATTTGAGTCCCC	2467
Qy	2549	AAACTACCCAGGCAATTATCCCAAGCCAAACACCAGTGTAGCTGGAGCATCAACCCACCCC	2608

Db 977 GGGCTGTGACCAATATTTGGCCGCAACACACAGTGGCGCAGCTTCGAATGCAGTTGCAAGAAAGG 1036
Qy 1136 ATTTAAATTTAATTAACAGATGAAAGCTTTGCCAAGATGTGGATGAGTCTCTTTGGATAG 1195
Db 1037 CTATAGCTTCTCAATAGAGAGAACTGGCCAGGATATAGACGAGTGTCTCTTTGATCG 1096
Qy 1196 GACCTGTGACCAAGCTGATCAACACACCCCTGGCACATTTGGTTGTGTGCAACCGAGG 1255
Db 1097 AACCTGTGACCAACATATGTGTCAACACACAGAGCTTCCAGTGTCTCTGCCATCGTGG 1156
Qy 1256 GTACACCTGTATGCTTACCCACTGTGAGACACCAATAGTCCAGCATCAACACGG 1315
Db 1157 CTACCTGTGTATGATATACCCACTGTGGGATGTGGATGAATGACGATCAACCGGGG 1216
Qy 1316 AGGCTGTGACGAGTGTGTGAACACAGTGGCGAGCTATGAATGCCAGTGGCACCCCTGG 1375
Db 1217 AGTTGGCCGCTTTGGCTGCATCAACACTCTGGCAGCTTACCAGTGTACTTGGCCAGAGG 1276
Qy 1376 ---GTAAAGCTTCACTGGAATTAATAAGACTGTGTGGAAGTGAAGGGCTCTCTGCCAC 1432
Db 1277 CCAGGTCGGCTGCATCGAATGGCAAGATTGGCAGAGCCACTGAAGTGTCAAGGGCAG 1336
Qy 1433 AAGTGTGTACCCCGTGTCTCCCTGCACTGCGGTAAAGTGTGTGAGAGAGCGGTGCTT 1492
Db 1337 TCTGGGGCTCGAAAGCATGCTCAGTGCACACCGGTCTGGCAAGAGGACACCTGTGC 1396
Qy 1493 CTTGAGATGTCACTCTGGCATTCACTCTCTTTCAGATGTCACCA-----CATCAGGAC 1546
Db 1397 CTTGACCTGTCCCTCAGGGCCGATTTTGGCAGAGGCTGCAGTGTCTCATTAACA 1456
Qy 1547 AAGTGTAACTTTAAGCTTAATTAAGGCAAGTGTAGTTGAA-----AATGTGAGCT 1600
Db 1457 ACGGGCTCTCTCAAGATCAAGATGCAAGTCAAGTCAAGTCTTGGACCTTGGCAAGAAAGGCAA 1516
Qy 1601 GTTTCCTGGAGGTCTCGGACAGGACTTACAGAGAGCACACTCAGTAAAGAGAGCTT 1660
Db 1517 AACAGAGGAGGTGGCAGAACACAGGCCAGGTGTGCTGCTCTCTGATGTCAGGT 1576
Qy 1661 CCGCTACGTAACCTTACATGAGCTCTGGCAAGCAAGTCCAGAGAGCCCTTGGCCGACC 1720
Db 1577 CACCTTCATCCACTTAAGTGTGACTCTCTCGAAGGCGCAGGCCGCGCCGAC 1636
Qy 1721 AAGACCCCTAAGGAAATGTTTA---TCACTGTTGAGTTTGAAGCTTGAACCTAACAATA 1777
Db 1637 CCTTCAGGCAAGAGGTCAAGAGGTCACTCCCTGGAACCTGGAGGCGAGGTCAAGAGCGA 1696
Qy 1778 GGAGGTGACAGCTCTTGTGACTGAGCTGCATCGTAAAGCGAACCGAGAGCGGCTCG 1837
Db 1697 AGAACCCAGCCAGCTGTGGGTGCTGCTCGCTCGACAGCGAATGGAAACGGCGGCTGAA 1756
Qy 1838 TAAAGCCATCCGACGCTCAGAAAGCCGCTCCACAGGAGAGTTCACCTCCAGCTCTC 1897
Db 1757 AGGATCCCTGAAGTCTCAGAAAGTCCATCAACAGGACCGCTTCTGCTGCGCTGCG 1816
Qy 1898 AGGCATGAACCTCGAGCTGGCTTAAAGGCTCCAGAACATCTGAACCGCAGGGA----- 1952
Db 1817 AGGCTCTGATATGAGCTGGCCCAAGCGCGGCTGTAGCGGGAGCGAGCAGAGGCC 1876
Qy 1953 ---GAGTCTGTGGAGTGGCGCAGGCTCATGCAGAAACCAATGTGTAGTGTGAGGGC 2008
Db 1877 GATGGAGTCTGTAGGCCCGGCGAGGACCGTGTGGGACCAAGTGTGTGAGCTGCCGCA 1936
Qy 2009 TGGGACCTATTATGATGGAGCAGAGACGCTGCATTTTATGTCCAAATGGAACCTTCCA 2068
Db 1937 GGGAACTATTACCGCGCCAGCAGCGAGCAGTGTGTGCCATGCCAGCGGGCACCTTCCA 1996
Qy 2069 AATAGGAGGAGCAATAGCTTGTGAACCATGCCCAACAGCAAGAAATCTGGGGGCTT 2128
Db 1997 GGAGAGAGAGGAGCAGCTCTCTCGCAGACCTTGGCTGGGAG-----TGATGCCCA 2047
Qy 2129 GAAGACCCAGAGCTTGGATATGTCTGAATGTGGAGGTCTGTCAACCTGGTGAATA 2188

Db 2248 CGGCGCTCTTTGGAGCCACCAAGCTCACCAGTGTGAGGTGAGTGGCCACCTGGCCAACA 2107
Qy 2189 TTCTGAGATGGCTTTGCACTTGGCAGCTCTGTGCTTGGCCTGGGACACTTCCAGCTGAAC 2248
Db 2108 CTCTGTAGATGGCTTCAAGCTTGTGAGCATGCCCACCTGSCACCTACCACTGAGC 2167
Qy 2249 TGGTCAACTTCTGCTTCCCTTGTGGAGGAGCTTGGCAACCAACATCAGGAGCTAC 2308
Db 2168 AGACAGAGCTTATGCTTCCCTTGTGGGGGCTTCCACCAACCAAGCTGAAGGGGCT 2227
Qy 2309 TTCTTTTCAAGTGTGAACAGAGATTCATGTTTCACTGTGACATTTCTAGACACCA 2368
Db 2228 TTCTTTTCAAGTGTGAACCAAGTTCAGTGTCTCCAGGAGCTTCTCCAGGAGCTTCTAGACACCA 2287
Qy 2369 CACTTCAACCTGATTTCTGCTTGGCCAGTGGAAACATACAGGCTGAATTTGGAAAAAATAA 2428
Db 2288 CATCCACCGCTGTATTTCTGCTGTGCCATGGCTCTTATCAGCCGACTTCCGTCAGACTT 2347
Qy 2429 TTGTGTTTCTTGGCCAGGAATACTAGGACTGACTTTGTATGGCTTCCACAAACATTAACCA 2488
Db 2348 CTGAGCGCTGTCCAGGAAACACAAAGCACAGACTTTGATGGCTCTTACCACTGTGTGCCCA 2407
Qy 2489 GTGTAAACACAGAGATGTGGAGGAGCTGGAGATTTTCACTGGGTACATTTGAATCCCT 2548
Db 2408 ATGCAAGATCTGTCAGTGTGGGAGCTGGGTGAGTTTCACTGGCTATTTGAGTCCCT 2467
Qy 2549 AAATCTACCGAGCAATTTAGCCAGCCAAACACCGAGTGTAGTGGACCATCAACCCACCCCT 2608
Db 2468 CAATCTACCGGCAACTACCCAGCTGGTGGAGTGCATCTGGAACATCAACCCCGCAGC 2527
Qy 2609 CAAGCGCGGATCTCTGATGTGTCTCTGAGATCTTCTTCTTGGCCATAGAGAGAGCTGTGG 2668
Db 2528 CAAGCGCAAGATCTTATCTGTGTACAGAGATCTTCTTCTTGGCCATCTGAGGATGAGT 2587
Qy 2669 GGACTATCTGTGTATCGGAAACCTCTTCACTCAATTTCTGTGACAAACATATGAAACCTG 2728
Db 2588 GGAGCTCTGTGTATGAGAAAGCACTTATCCCATCTCTTATACACTTATGAGACCTG 2647
Qy 2729 CCAGACTACGAACCGCCCATCGCTTCACTTCCAGGTCAAGAGAGCTGTGGATTCAGTT 2788
Db 2648 CCAGACTACGAGCGTCCCATTTGCTTCACTGCTTCCAGGAGCTTCCAGAGCTCTGGATCAACT 2707
Qy 2789 CAAGTCAATGAAGGAAACAGCTTACAGGTTCCAGGTTCCATACGTGACATATGATGA 2848
Db 2708 CAAGCAAGCGAGGCAACAGCGCTGCTTCCAGTCTTCCAGTCTGATGA 2767
Qy 2849 GSACTACCGAACTCATTTGAAGACATAGTGGAGATGGCAGGCTCTGATGCACTGAGAA 2908
Db 2768 GSACTATGAGCAGCTGTGAGAGACATTTGTGAGATGGCGGCTCTATGCTCTGAAA 2827
Qy 2909 CCATCAGGAAATACTTAAAGATAAGAACTTATCAAGGCTCTTGTGATCTCTTGGCCCA 2968
Db 2828 CCACCGAGGATTTTAAAGCAAGAGCTCATCAAGGCTTCTTGTAGGTTGCTAGCCCA 2887
Qy 2969 TCCCGAGAACTTTCAAGTACACAGCCAGGAGTCCCGAGAGATGTTTCCAGATCGTT 3028
Db 2888 CCCCCAGAACTTCAAGTACACAG---AGAAACACAAAGGAGATCTGCCAAAATCTTT 2944
Qy 3029 CATCCGATCTGCTAGCTTCAAGTGTCCAGGTTTTCAGAGCTTACAAA 3078
Db 2945 CATCAAGCTGCTCGCTCCAAAGTTTCCAGCTTCTTGAGGCCCTACAAAT 2994

RESULT 10

US-0-95-142-1
; Sequence : Application US/10:95142
; Publication No. US20010036143A
; GENERAL INFORMATION:
; APPLICANT: Myriad Genetics, Inc.
; APPLICANT: Wettstein, Daniel; Albert
; APPLICANT: Mauck, Kimberly A
; TITLE OF INVENTION: NOVEL PN826 NUCLEIC ACIDS AND USE THEREOF
; FILE REFERENCE: 1051.01

```

: CURRENT APPLICATION NUMBER: US/10/195,142
:
: CURRENT FILING DATE: 2002-07-10
:
: PRIOR APPLICATION NUMBER: US 60/304,323
:
: PRIOR FILING DATE: 2001-07-10
:
: NUMBER OF SEQ ID NOS: 39
:
: SOFTWARE: PatentIn version 3.1
:
: SEQ ID NO 1
:
: LENGTH: 3791
:
: TYPE: DNA
:
: ORGANISM: Homo sapiens
:
: FEATURE:
:
: NAME/KEY: CDS
:
: LOCATION: {52}..{3018}
:
: OTHER INFORMATION:
:
: US-10-195-142-1

```

[illegible]

Db	774	----	773
Qy	926	CATGGAAACGTGTGCTGTCAACAATGAGGCTGTGACCGCACTGTGAAGATACCTTCGAC	985
Db	774		833
Qy	986	AGGTGTCCTACTGCACTGTCTCTGTTGGATTCACCTCTCCAGTTGGATGGGAACACATGTAA	1045
Db	834	TGCGCTGCGATGCACTGCCCGCTTGATTCACATGTCAGCGCGGACGGGAACACATGCAA	893
Qy	1046	AGATATTGTAGCTGCGACACCGCAATCGAGGTTGTGATCATTTCTGCAAAACATCGT	1105
Db	894	AGACATCAACGAGTGCTGGTCAAACACGGAGGCTCGACCACTTCTGCGCAACACCGT	953
Qy	1106	GGSCAGTTTGACTGGCGCTGCGAAGAAGATTTAAATATTAAACAGATGAGAAGTCTTG	1165
Db	954	GGSCAGTTTGACTGGCGCTGCGAAGAAGGCTACAAGCTGCTACCGGACGAGCGCACTG	1013
Qy	1166	CGAAGATGGATGAGTGTCTTTTGGATAGCACTGTGACCAAGCTGTGATCAACACCC	1222
Db	1014	CGAGGACATGGACGAGTGCTCTTTCAGCGGACCTGTGACCACTCTGATCAACTCCCC	1073
Qy	1226	TGGCACATTTGCTTGTGCTTGCACCGAGGATACACCTGTATGCGTTTCAACCACCTGTGG	1285
Db	1074	GGCAGCTTCCAGTGCCTGTGTG:CACCGCGCTACATCTCTACGGACAAACCCACTCGG	1133
Qy	1286	AGACACAAATGAGTGCAGCATCAACAAACGGAGGCTGTACGACAGGCTGTGTGAAACAGT	1345
Db	1134	AGATGTGGACGAGTGCAGCATGACAAACGGAGCTGTGACCAAGGCTGTGTCAACACCAA	1193
Qy	1346	GGCAGCTATGAAATGCACTGCCACCTGGGTACAAGCTCGACTGGAATAAAAGACCTG	1405
Db	1194	GGCAGCTATGAGTGTGTCTGTCCCCCGGAGGCGGCTCCACTGAAACGGGAAGATTTG	1253
Qy	1406	TGTGGA---AGTCAAAGGGGCTCTGGCCCAAGTGTGTACCCCGTGTCTCCCTGCAGTG	1462
Db	1254	CGTGGAGACAGGCAAGTGTCTTTCTCGGSCAAGACCTCCCGCCGGCCAGCTGTCTCTG	1313
Qy	1463	CGGTACAGTGTGTGGAGAGACGGGTCTCTCTCATGTCACTCTTGGAATTACCACTCTC	1522
Db	1314	CAGCAAGGCAGGCGGTGTGGAGAGCTGCTCTCTTCTGCCCGGCTCACACACTCTTCGT	1373
Qy	1523	TTCAAGT-----	1530
Db	1374	GCACAGCTCGAATAATAGCTACGCTCGAGCTCGGAGTTCCAGGCGCCAGCGGCAAGGC	1433
Qy	1531	-----TCACCAACCATCAGGACAAAGTGTATACCTT	1555
Db	1434	GCTGCAAGAACGCAACGGCACCAAGCTGTGCGCTCGGCGCCAGCTGTGTAGATCCGCCAC	1493
Qy	1559	TACGCTAAATGAAGGCAAGTGTAGTTGAA-----AAATGCTAGCTGTTTCCCGAGGG	1612
Db	1494	CACCCCATCAAAACAGAAAGCCCGCTTCAAGATCCGAGATGCCAGTGCACCTCTCGGCC	1553
Qy	1613	TCTGCGACCAAGCACTACAGAGAAGCACAGCT-----CAGTAAAGAGAGCTTCCGCTA	1666
Db	1554	CCACAGCCAGCAGCAAGAGAGACCGCCAGGACCGCTCTGGACCACTGCCATGT	1613
Qy	1667	CGTAAACCTTACATGAGCTCTGGGAAGCAGTCCAGGAGCCCTTGGCGGACCAAGCAC	1726
Db	1614	GACTTTGCTGACCTCAAGTGTACTCTCTCAAGAGAGGCGCGGTGGCGCAAGTCCCC	1673
Qy	1727	CCGTAGGAAATG---TTATCACTGTTGAGTTTGAATCTAAACCAAGAGAGGT	1783
Db	1674	ATCCAGGAGGTGTCCACATCACAGCAGTTTGTGATCGACAAAGATGGAAGAGC	1733
Qy	1784	GACAGCTTCTTGTAAGCTGAGCTGCATGATAAGCGAACCGGATCCCTTAAAGC	1843
Db	1734	CTCAGACATGGAAGCGACTGCTTGGCGAAGCGAGACAGAGCTGAGGCGCC	1793
Qy	1844	CATCCGACGCTCAGAAAGGCGCTCCACAGGAGAGTTTCACTCTCAGCTCTCAGGCAT	1903


```

Db 2706 GCGCATGCGCTTCACTCCGCTCCGCAAGCTCTGGATCCAGTTCAAAATCAATGAAGG 2765
Qy 2804 GAACAGCGCTAGAGGCTTCCAGGTCCTCCATACGTGACATATGATGAGGACTACCCAGGAACT 2863
Db 2766 CAACAGCGCAAGGCTTCCAAAGTCCCTA.TGTACCTACGATGAGGACTACCCAGCAACT 2825
Qy 2864 CATTAAGACATAGTTTCGAGATGGGAGGCTCTATGCACTGAGAACCAATCAGGAAATACT 2923
Db 2826 CATAGAGACATCGTGGCGGATGGGCGCTGTACGCCCTGGAGAACCAACCAAGGAAATTT 2885
Qy 2924 TAAGGATAGAAATATTAAGGCTCTGTTCATGCTCTGCGCCATCCCGAGACTATTT 2983
Db 2886 GAAAGACAAGAGCTGTATCAAGGCCCTCTTCAGCGTGTGGCGCATCCCGAGACTACTT 2945
Qy 2984 CAAGTACACAGCCAGGAGTCCCGAGAGATGTTTCCAGATCGTTCATCCGATTCCTACG 3043
Db 2946 CAAGTACACAGCCAGGATCCCAAGGAGATGTTCCACAGGCTCTTCATCAAACTGCTCG 3005
Qy 3044 TTCCAAAGTGTCCAGTGTTCAGACCTTTTACAAATGAC 3081
Db 3006 CTCCAAAGTGTCTCGTCTCTCGGCCCTTACAAATAAC 3043

RESULT 12
US-10-195-142-3
; Sequence 3, Application US/10195142
; Publication No. US20030036163A1
; GENERAL INFORMATION:
; APPLICANT: Myriad Genetics, Inc.
; APPLICANT: Wettstein, Daniel Albert
; APPLICANT: Mauck, Kimberly A
; TITLE OF INVENTION: NOVEL PN9826 NUCLEIC ACIDS AND USE THEREOF
; FILE REFERENCE: 1051.01
; CURRENT APPLICATION NUMBER: US/10/195,142
; CURRENT FILING DATE: 2002-07-10
; PRIOR APPLICATION NUMBER: US 60/304,323
; PRIOR FILING DATE: 2001-07-10
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 3
; LENGTH: 2967
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-195-142-3

Query Match 28.2%; Score 1052.2; DB 14; Length 2967;
Best Local Similarity 62.7%; P-red No 5e-311;
Matches 1903; Conservative 0; Mismatches 923; Indels 207; Gaps 9;

Qy 153 CTGCTGCTGCTGCGGGGCGCTCCGCGGGTGGGGCGCGTCCCGGGGCGGAGGAG 212
Db 37 CTGCTGGCCCTGGGCACACGCGGGGCGCTGGCGGGGCGAGCGGGTCCCGAGGTCAGTC 96
Qy 213 GATGTAGATGAGTGTCCCAAGGCTAGATGACTGCATGCCAGCGCCCTGTGTAGAAC 272
Db 97 GACGTGGATGAGTGTCTCAGAGGCGCAGATGACTGCCACATCATGCCCATCTGTAGAAC 156
Qy 273 ACACCCACCTCTACAAAGTGTCTCTCAAGCTGGCTACCAAGGGAAGGAGGAGTGT 332
Db 157 ACGCCCAAGTCTTCAAAATGCTCTCTCAAGCAGGCTACCAAGGGAAGGAGGAGTGT 216
Qy 333 GAGGACATCGATGAATGTGGAATG---AGCTCAATGGAGGCTGTGTCCATGACTGTITG 389
Db 217 GAAGACATTGACAGTGTGAGATGACTACTCAATGGGGGCTGTGTCCACAGTGCATC 276
Qy 390 AATAATTCAGGCAATATGTTGCACTTGTGTTGATGGCTTATGTTGGCTCATACAGGT 449
Db 277 AACATCCCGGGAACTACAGAGGTACCTGCTTGTATGGCTTCATGTGGCACACATGGA 336
Qy 450 CATAATTTGCTTGTATGTGACAGTGTCTGGAGAACATGGCGGCTGGCAGCATACCTGT 509
Db 337 CACAACCTGCTGTGATGTGGACGAGTGTCCAGGCAATAATGGTGGCTGCCAGGAGATCTGC 396

```

```

Qy 510 GTCAACGTGATGGGAGCTATGAGTGTCTGTGCAAGAGAGGGTGTTCCTCTGAGTGACAT 569
Db 397 GTCAATGCCATGGGAGCTACGAGTGTCTGAGTGTGCAAGAGAGGGTGTTCCTCTGAGTGA 456
Qy 570 CAGGACACCTGCAATTCACCGCTCGGAGAGGGCTTGAGCTGCATGAATAGGAGTCAAGGC 629
Db 457 CAGCATACCTGCAATTCACCGCTCGGAGAGGGCTTGAGCTGCATGAATAGGAGTCAAGGC 516
Qy 630 TGTAGTCACTCTGCAAGAGGAGGCTTCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 689
Db 517 TGTGCTCACTCTGCAAGAGGAGGCTTCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 576
Qy 690 TTTAGTGTGCTGCAAGAGGAGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 749
Db 577 TTTAGCTTGTGCAAGAGGAGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 636
Qy 750 TGGCAGCAGTCTGCTGCAAGAGGAGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 809
Db 637 TGGCAGCAGTCTGCTGCAAGAGGAGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 686
Qy 810 AAGATGCAACAGTGTGCTGCAAGAGGAGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTG 869
Db 697 GCGCTCCACTCAGACGCTGCAAGAGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 722
Qy 870 GAGAGCAACACCATCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 929
Db 723 -----CATC 726
Qy 930 GMAAGCTGTCTGTCAACAAATGGAGGCTGTGAGCGGACCTGTAAAGGATATCTGACAGGT 989
Db 727 GAGAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 786
Qy 990 GTCCACTGCACTTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1049
Db 787 GTGCGATGCACTTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 846
Qy 1050 ATTGATGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1109
Db 847 ATCAACGAGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 906
Qy 1110 AGTTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1169
Db 907 AGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 966
Qy 1170 GATGTGATGAGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1223
Db 967 GACATGCAAGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1026
Qy 1230 ACATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1289
Db 1027 AGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1086
Qy 1290 ACCAATGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1349
Db 1027 GTGAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1146
Qy 1350 AGCTATGAATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1409
Db 1147 AGCTACGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1206
Qy 1410 GA----AGTGAAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1466
Db 1207 GAGACAGGCAAGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1266
Qy 1467 AAGAGTGTGAGAGAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1526
Db 1267 AAGCAGGCGGCTGTGAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1326
Qy 1527 GA----- 1528
Db 1327 GACTCGGAAAATAGCTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1386

```

Qy	1529	- - - - - TGT CACCACCATCAGGACAAGTGTAACTTTAAG	1566
Db	1387	CAGAAACGCAACGGCACCAAGCTCTGGGCGCCAGATGTCCTCAGATGCCCCCAGCACAC	1446
Qy	1563	CTAAATGAAGGCAGTGTTAGTTTAA- - - - - AAATGCTGAGCTGTTTTCCGAGGAGTCTG	1616
Db	1447	CCCATCAAACAGAAGGCCCTCTTAAGATCCGAGATGCCAAGTGGCACCTCTCCGGCCCCAC	1506
Qy	1617	CGACCAGCACTACACAGAAAGCACAGCT- - - - - CAGTAAAGAGAGACTTCCCCTTACGTA	1670
Db	1507	AGCCAGGCACGACAAGAGAGACCGCCAGCAGCCGGTCTGGACCACTGCCATGTGACT	1565
Qy	1671	AACCTTACATGCAGCTCTGCGAAGCAAGTCCCAGAGGCCCTGGCGGACCAAGSACCCCT	1730
Db	1567	TTCGTGACCCCTCAAGTGTGACTCTCTCAAGAGAGGGCGCTGGCGCAAGTCCCCATCTC	1626
Qy	1731	AAGGMAATGT- - - TTATCACTGTTGAGTTTGAGCTTTGAAACTAACCAAGGAGGTTGACA	1787
Db	1627	AAGGAGTGTCCCACTACACAGCAGAGTTTGAATCGAGACAAGATGGAAGAGGGCCTCA	1688
Qy	1789	GCTTCTTTGACCTGASCTGCATCGTTAAAGCGAACCGAGAGCGGCTCCCTAAAGGCATC	1847
Db	1687	GACACATGCCNAGCGGACTGCTTCGGNAAGGACAGACAGAGCCTGCAGGCCGCATC	1746
Qy	1848	CGACGCTCAGAAGAGCGCTGCCACAGGGAGAGTTTCACCTCCAGCTCTCAGGCACTGAAC	1907
Db	1747	AAGACCCCTGGCAAGTCCATCGGCCGACAGTTCTATGCTCCAGGCTCTCAGGCACTGAG	1806
Qy	1908	CTCGAGCTGGCTAAAAGACCTCCACAGAACATCTGAAGCCAGCAGAGTCTCTGGAGTG	1967
Db	1807	TACGAGGTAGCCACAGGCCAGCCNAGGCGCTGGAGGGGAG- - - GGGGCACTGGGCGCA	1863
Qy	1968	GGCCAGGGTTCATGCAGAAAAACCAATGTGTGAGTTGCAAGGCTGGGAGCTATTATGATGGA	2027
Db	1864	GGCCAGCTGCTACAGGACAGCAATGCGTTGCCCTGTGGGCTGGCACCCACTTCGGTGGT	1923
Qy	2028	GCACGAAACGCTGCATTTTATGTCCAATTGGAAACCTTCCAATGAGGAGGACAAATG	2087
Db	1924	GAGCTGGCCAGTGTGTGTATGTATGCCAGGAACATACCAGGACATGGAAGGCCAGCTC	1983
Qy	2088	ACTTGTGAACCATGCCCAAGACAGGAAATCTCTGGGCCCTGAAGACCCAGAGGCTGG	2147
Db	1984	AGTTGCAACCGTGCCCCAGCAGCGGGCTTGGTCTG- - - - - CTGGTGCCCGC	2034
Qy	2148	AAATATGTCTGAATCTGGAGGTCTCTGTCAACCTGGTGAATATCTGCAGATGGCTTGCAC	2207
Db	2035	AACGTGTGGNAATGTGGAGGCCAGTGTCTCCAGGCTCTCTCGGCCGATGCTTCAAG	2094
Qy	2208	CCTTGCCAGCTCTGTGGCTCGGGCAGCTTCCAGCCTGAAGCTGGTGGAACTTCTTGCTTC	2267
Db	2095	CCCTGCCAGGCTCGCCCGTGGGCACGTACCAGCCTGAGCCCGGGCGCACGGCTGCTTC	2154
Qy	2268	CCCTGTGGAGGAGCCTTGGCCACCAACATCAGGAGCTACTTCTCTTCAGGACTGTGA	2327
Db	2155	CCCCTGTGAGGGGGTTTGCTCACCANAACGNAAGCACCACTCTTCCAGGACTGGCAG	2214
Qy	2328	ACCAGAGTTCAATGTTCACTCTGACATTTTCTACAAACCAACCACTCACCCGATGTATTCGT	2387
Db	2215	GCTAAAGTGCACTGCTCCCCGGCCACCCTACAAACCAACCAACCCGCTGCATCCCG	2274
Qy	2388	TGCCAGTGGGAACATACACAGCTGAATTTGGAAAAATAATATGTGTCTTCTGCCCCAGGA	2447
Db	2275	TGCCCCGTGGCACCTTACCAAGCCGAGTTTGGCCAGAACCACTGCATCACTGTGCCGGGC	2334
Qy	2448	AATACTAGCTGACTTGTATGGCTCCCAACACATACCCAGTGTAAACACAGAGATGT	2507
Db	2335	AACACCAACAGACTTCGATGGGTCCCAACAGCTCACACTGCAAAAACCAACCACTGC	2394
Qy	2508	GGAGGGAGCTGGGAGATTTTCACTGGGTACATTTGAATCCCCCAACTACCCAGCAATTC	2567
Db	2395	GGCGCGAGCTTGGTGACTACACGGGTATCATGAGTGTCCCACTACCTGCGGCACTAC	2454
Qy	2568	CCAGCCAAACCGAGTGTTACGTGGACCACTCAACCCCAAGCGCGCATCTCTGATC	2627

DB	2455	CCAGCCAAAGCCTCAATGGCGTCTGGCACATCGCACCTCCCGCCCAAGGCGAGGATCCTCATC	2514
QY	2628	GTGGTCCCTCAGATCTTCCCTGCCCATAGAGAGCACTGTGGGGACTATCTTGGTGAATGGCG	2687
DB	2515	GTGGTCCCTCAGATCTTCCCTGCCCATCGAGGATGAGTGGCGGATGCTTCTGTCATGAGG	2574
QY	2688	AAAACCTCTTTCATCCAAATCTGTGACAAACATATGAAACCTGGCAGACCTACGAGACGCCCC	2747
DB	2575	AAGAGTGCCCTCTCCACGCTCCATCACACCTATGAGACCTGSCAGACCTACGAGAGGCCCC	2634
QY	2748	ATGGCCTTCACCTCCAGGTCMAAGAGAGCTGTGGATTCAGTTCAGTCCAGTCCCATGAAGGGAAC	2807
DB	2635	ATCGCCTTCACCTCCCGCTCCCGCAAGCTCTGATCCAGTTTCAAAATCCCAATGAAGGCAAC	2694
QY	2808	AGCGCTAGAGGGTCCAGAGTCCCATACGTGACATATGATGAGCACTACCAGGAACCTCAT	2867
DB	2695	AGGGCAAGAGCTTCCAGTGCCCTATGTACACTAGTAGGAGACTACGACCACTCAT	2754
QY	2868	GAAGACATAGTTCGAGATGGCAGGCTCTATGCATCTGAGAACCATCAGSAATACTTTAAG	2927
DB	2755	GAGGACATCTGCGGATGGGCGCTGTACGCCCTCGAGAAACCCAGGAAATTTTGA	2814
QY	2928	GATAGAAATTTACAGGCTCTGTGATGTCCTGGCCATCCCGAGACCTATTTCAAG	2987
DB	2815	GACAAGAGCTGATCAAGGCCCTCTTCGAGCTGCTGGCGCATCCCGAGAACTACTTCAAG	2874
QY	2988	TACACGCCCAGAGTCCCGAGAGATGTTTCCAAAGATCGTTTCATCCGATTCTACGTTCC	3047
DB	2875	TACACAGCCAGAAATCCAGAGGAGATGTTCCCGAGGATCTCTCATCAACATGCTCGCGTCC	2934
QY	3048	AAAGTGTCAGGTTTGTGAGACCTTCAAAATGA	3120
DB	2935	AAAGTGCTCGGTTCTCGCGGCTTCAAAATAA	2967

```

RESULT 13
US-C9-899-570-15
: Sequence 15, Application US/0598570
: Patent No. US20020123612A)
: GENERAL INFORMATION:
: APPLICANT: GERLACH, VALERIE L.
: APPLICANT: ELLERMAN, KAREN
: APPLICANT: MACCOUGALL, JOHN R.
: APPLICANT: SMITHSON, GLENDA
: TITLE OF INVENTION: NOVEL HUMAN PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND
: TITLE OF INVENTION: METHODS OF USING THE SAME
: FILE REFERENCE: 15966-776CIP
: CURRENT APPLICATION NUMBER: US/09/998,570
: CURRENT FILING DATE: 2001-07-03
: PRIOR APPLICATION NUMBER: 60/198,293
: PRIOR FILING DATE: 2000-04-19
: PRIOR APPLICATION NUMBER: 60/198,645
: PRIOR FILING DATE: 2000-04-20
: PRIOR APPLICATION NUMBER: 60/210,809
: PRIOR FILING DATE: 2000-06-09
: PRIOR APPLICATION NUMBER: 60/193,476
: PRIOR FILING DATE: 2000-04-26
: PRIOR APPLICATION NUMBER: 60/200,025
: PRIOR FILING DATE: 2000-04-26
: PRIOR APPLICATION NUMBER: 60/224,610
: PRIOR FILING DATE: 2000-08-11
: PRIOR APPLICATION NUMBER: 60/200,024
: PRIOR FILING DATE: 2000-04-26
: PRIOR APPLICATION NUMBER: 60/193,880
: PRIOR FILING DATE: 2000-04-26
: PRIOR APPLICATION NUMBER: 60/228,591
: PRIOR FILING DATE: 2000-07-17
: PRIOR APPLICATION NUMBER: 60/271,814
: PRIOR FILING DATE: 2001-02-27
: PRIOR APPLICATION NUMBER: 60/215,855
: PRIOR FILING DATE: 2000-07-03
: PRIOR APPLICATION NUMBER: 09/833,446

```

QY	176	CCGCGCGGTCGGGCGGTGCGCGGGCGCGAGGAGTGTAGATGAGTGTGCCCCAGG	233
Db	149	CAGCGCGCGCGCGAGTAGCAAAAGCGC-GCAAGATGTGTGATGAGTGTGTGGAGGG	207
QY	236	GCTAGATGACTGGCATGCCGACGCCCTGTGT-CAGAACACACCCACCTCTACAAGTGCTC	295
Db	209	GACTGACAACTGCCACATCGATGCTA-CTGCCAGAACACCCGAGGTCA-ACAAGTGCAT	267
QY	296	CTCGAAGCTGGCTACCAAGGGGAAGGAGCGCTGTGAGAGCATCGATGAAATGTGGAAA	355
Db	268	CTGCAAGTCTGGCTACACAGGGGACGGCAACACTGCAAGACGCTGGATGAGTGCAGAGC	327
QY	356	TGAGCTCAATGGAGGCTGTCTCCATGACTGTTTGAATATTCAGGCAATATTCGTTGCAC	415
Db	328	AGAGGATAATGCAAGTTGTGTGCATGACTGTGTCAACATCCCTGGCAATTAACCGGTGAC	387
QY	416	TTGTTTTGATGGCTTTCATGTGGCTCATGACGGTCATAAATGTCTTGTATGTGGACGAGTG	475
Db	388	CTGCTATGATGATTCACCTGGCACATGACGGACACAAC-TGCTGGATGTGGACGAGTG	447
QY	476	CTTGGAGAACAAATGGCGGCTCCAGCATACCTGTGTCAAGCTCATGGGAGCTATGAGTG	535
Db	448	TGCCGAGGGCAACGGCGCTGTACGACAGCTGTGTCAACATGATGGCAGCTATGAGTG	507
QY	536	CTGCTGCAAGGAGGGTTTTCTCTGAGTGTGAATCAGCACACCTGCTATTCACCGCTCGGA	595
Db	508	CCACTGCGGGAAGGCTTCTTCTCAGCGAACGAGCATACCT-GTA-TCCAGCGGCGCAGA	567
QY	596	AGAGGGCTTGAGCTGCATGAATAAGGATCACGGCTGTAGTCACAT-CTCAAGSAGGCGCC	655
Db	568	AGAAGGAATGAATTTGCATGAACAGAACACCGCTGTGCCACAT-TTGCCGGAGACACC	627
QY	656	AAGGGGAGCGGTGCGCTGTGAGTGACGCGCTGGTTTTGAGTGTGCCAAGAACCGAGAGA	715
Db	628	CAAGGGGGGTATTGCTGTGAATGCCGTCTGGCTTTGAGCTTACCAAGAACCAACGGGA	687
QY	716	CTGCATCTTGACCTGTAAACCATGGGAAGCGGTGGGTGCCAGACCT-CCTGTGACGATACGC	775
Db	688	CTGTAAATTGACATGCAACTATGTATAACGGCGGTGCCAGACACAGTGTGATGACACAGA	747
QY	776	CGATGGCCCCAGAGTGCAGCTGCCATCCACGTACAAGATGCACACAGATGGAGGAGCTG	835
Db	748	GCAGGGTCCCCGTGCGGCTGCCATATCAAGTTTGTGCTCCATACCGACGCGGAACAGATG	807
QY	836	CCTTGAGCGAGGAGACCTGTCTCTGGAGGTGACAGAGAGCAACACCATCAGTGGTGGGA	895
Db	808	CATCGGG-----GAAAGGGGGCTAGAGCAGCACATCCCCACT-----	844
QY	896	TGGGGATAAACGGGTGAACCGCGCGCTGCTCATGGAAACGTGTGCTGTCAACAATCGAGG	955
Db	845	-----CAAGCCGTTTCTAATGAGACCTGTGCTGTCAACAACCGGGG	885
QY	956	CTGTGACCGCACCTGTAGGATACCTTCGACAGGTGTCCACTGCACTTGTCTGTTGGATT	1015
Db	886	CTGTGACACAGTAAGTGCCCATGATGACGACACTGGTGTCCACTGCACCTCCCTCTGTGGGCTT	945
QY	1016	CACCTCTCCAGTTGGATGGGAAGACATGTAAAGATATTGATGAGTGCCAGACCGCAATGG	1075

Db 2026 GCCCGGACGACCGTGTCTGGGACCAAGTGTGTAGCTGCCCGCAGGAAAGTATTACCA 2085
 QY 2024 TGAGACACGAGACCGTGCATTTTATGTCCAAATGGACCTTCCAAATGAGGAGGACA 2083
 Db 2086 CGGCCACAGACGAGACGAGTGTGTCCATGCCACCGGCGACCTTCCAGAGAGAGAGGACA 2145
 QY 2084 AATGACTTGTGAACCATGCCACAGACAGGAAATCTGGGGCCCTGAAGACGCCAGAGAGC 2143
 Db 2146 GCTCTCTCGACCTTGGCCCTGGAG-----TGTGCCACGGCCCTTTGGAGC 2196
 QY 2144 TTGGAAATAGTCTGAATGTGAGGCTGTGTCAACTGTGTGAATATTCGAGATGGCTT 2203
 Db 2197 CACCAACGTCACACCGTGTGACGAGTCACTGCCACCTGCTGCCAACACTCTGTAGATGGCTT 2256
 QY 2204 TGCACCTTCCGAGCTGTGTGCTGCGCCTGGSCAGCTTCCAGCTGAAGCTGTGGAACTTCTCTG 2263
 Db 2257 CAAGCCCTGTAGCCATGCCACGTGGCACCTTACCAACCTGAAGCAGGACGACCCCTATG 2316
 QY 2264 CTTCCCTGTGGAGGAGGCTTGGCCACCAAAATCAGGAGGCTACTTCTTTTCAGAGCTG 2323
 Db 2317 CTTCCCTGTGGTGGGGCTTCCACCAACGATGAGGGCCATTCTTCCCAAGACTG 2376
 QY 2324 TGAACACAGGTTCAATGTTCACCTGGACATTTCTTACACACACCACTTCCCGATGTAT 2383
 Db 2377 TGACACCAAGTCCAGTGTCTCCCGAGGCACTACTACACACAGCATCCACCGCTGTAT 2436
 QY 2384 TCGTTGCCAGTGGGAACATACAGCTCAATTTGGNAAATAATTTGTGTTCTTGCCC 2443
 Db 2437 TCGTGTGCACTGGGCTCTTATCAGCCGACTTCCGTCAAGACTTCTGACGCGCTGTCC 2496
 QY 2444 AGGAATACTAGCACTGACTTTGATGGCTCCCAACACATAACCCAGTGTAAACAGAG 2503
 Db 2497 AGGAACACACACAGACTTTGATGGCTCTACAGTGTGGCCCAATGCAAGATCGTCA 2556
 QY 2504 ATGTGAGGGGAGCTGGAGATTTCACTGGGTACATTGATGATCCCAACTACCCAGGCAA 2563
 Db 2557 GTGTGTGGGAGCTGGGTGAGTTCACTGGCTATATTGAGTCCCCCACTACCCGGGCAA 2616
 QY 2564 TTACCCAGGCAACACCGAGTGTACGTGGACCACTCAACCCACCCCGCCGCGATCCT 2623
 Db 2617 CTACCCAGCTGTGTGGAGTGATCTGGAACTCAACCCCGCCAGCGCAAGATCCT 2676
 QY 2624 GATCGTGTCCCTGAGATCTTCTGCCCCATAGAGGACGACTGTGGGACTATCTGGTAT 2683
 Db 2677 TATCGTGTACACAGATCTTCTGCGCATCTCAGGATGAGTGTGGGACGTCCTGTCAT 2736
 QY 2684 GGGAAACCTCTTCATCAATCTGTGACACATATGAACCTGCGACGCTAGCAAG 2743
 Db 2737 GAGAAAGAACTCATCCCCATCTCCATTACCACTTAAGACCTGCGACGCTAGAGCG 2796
 QY 2744 CCCCATCGCCTTCACTCCAGGTCGAAGAGCTGTGGATTCAAGTCCAAATGAAG 2803
 Db 2797 TCCATTGCTTCACTGCGCTTCCAGGAAGCTCTGGATCAACTCAAGACAGCGAGC 2856
 QY 2804 GAACAGGCTAGAGGTTCCAGTCCCATACGTGACATATGATGAGGACTACCAAGACT 2863
 Db 2857 CAACAGCGCCGTGGCTTCCAGATTCCTATGTGTTACCTATGATGAGGACTATGACAGCT 2916
 QY 2864 CATTAAGACATAGTTCGAGATGGAGGCTTATGATCTGAGAACCATCAGGAAATCT 2923
 Db 2917 GGTAGAAGACATTTGCGAGATGGCGGCTTATGCTCTGAAACACCAAGAGATTTT 2976
 QY 2924 TAAGTAAGAAATTTTCAAGGCTCTGTTGATGTCTTGGCCCATCCCGAGACTATTT 2983
 Db 2977 AAAGGACAGAACTCATCAAGGCTTCTTTGAGGTGTAGCCACCCCGCCAGACTACTT 3036
 QY 2984 CAAGTACAGCCAGGAGTCCGAGAGATGTTTCCAAAGATGTTTCACTCCGATGCTAGC 3043
 Db 3037 CAAGTACAG---AGAAACACAAAGAGATGCTGCCAAAATCTTTCATCAAGCTCTCCG 3093
 QY 3044 TTCCAAAGTGTCCAGGTTTTTGAGACTTACAAAT 3078
 Db 3094 CTCCAAAGTTTCCAGGTTCTTGAGGCCCTTACAAAT 3128

RESULT 14

US-09-839-446-15
 ; Sequence 15, Application US/39839446
 ; Publication No. US20030050232A1
 ; GENERAL INFORMATION:
 ; APPLICANT: GELACH, VALERIE L.
 ; APPLICANT: ELLERMAN, KAREN
 ; APPLICANT: MACDOUGALL, JOHN R.
 ; APPLICANT: SMITHSON, GLENDA
 ; TITLE OF INVENTION: NOVEL HUMAN PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND
 ; FILE OF INVENTION: METHODS OF USING THE SAME
 ; FILE REFERENCE: 5965-776
 ; CURRENT APPLICATION NUMBER: US/09/839,446
 ; CURRENT FILING DATE: 2001-04-13
 ; PRIOR APPLICATION NUMBER: 60/199,293
 ; PRIOR FILING DATE: 2000-04-19
 ; PRIOR APPLICATION NUMBER: 60/198,645
 ; PRIOR FILING DATE: 2000-04-20
 ; PRIOR APPLICATION NUMBER: 60/200,809
 ; PRIOR FILING DATE: 2000-06-09
 ; PRIOR APPLICATION NUMBER: 60/199,476
 ; PRIOR FILING DATE: 2000-04-26
 ; PRIOR APPLICATION NUMBER: 60/200,025
 ; PRIOR FILING DATE: 2000-04-26
 ; PRIOR APPLICATION NUMBER: 60/224,610
 ; PRIOR FILING DATE: 2000-08-11
 ; PRIOR APPLICATION NUMBER: 60/200,024
 ; PRIOR FILING DATE: 2000-04-26
 ; PRIOR APPLICATION NUMBER: 60/199,980
 ; PRIOR FILING DATE: 2000-04-26
 ; PRIOR APPLICATION NUMBER: 60/218,591
 ; PRIOR FILING DATE: 2000-07-17
 ; PRIOR APPLICATION NUMBER: 60/271,814
 ; PRIOR FILING DATE: 2001-02-27
 ; NUMBER OF SEQ ID NOS: 49
 ; SOFTWARE: Patent In Ver. 2.1
 ; SEQ ID NO 15
 ; LENGTH: 3177
 ; TYPE: DNA
 ; ORGANISM: Unknown Organism
 ; FEATURE:
 ; OTHER INFORMATION: Description of Unknown Organism: POLYX
 ; OTHER INFORMATION: CG55098-04
 US-09-839-446-15

Query Match: 25.7%; Score 959.4; DB 11; Length 3177;
 Best Local Similarity 60.9%; Pred. No. 1.5e-282;
 Matches 1847; Conservative 0; Mismatches 1001; Indels 187; Gaps 10;
 QY 176 CCGCGCGGTCGGGGCGGTGCCCGGGCGGAGGAGGATGATGATGAGTGTGCCCAAGG 235
 Db 149 CAGCCCGCGCCCTCAGTACAGCAAGCGGC-GCAGATGTGGATGATGATGTGGAGG 207
 QY 236 GCTAGATGATGCTCCATCCGAGCCCTGTGTGAGAACACACCACTCCCTCAAGTGTCTC 295
 Db 208 GACTGCAACTGCCCATGCTATCTGCCAGAACACCCCGAGGTATACAGTGCAT 267
 QY 296 CTGCAAGCTGGTACCAAGGGGAAAGGAGGAGTGTGAGGACATCGATGAATGTGGAAA 355
 Db 268 CTGCAAGTCTGGGTACACAGGGGACGGCAACACTGCAAGAGCGTGGATGAGTGGAGCG 327
 QY 356 TGAGCTCAATGGAGGCTGTGCCATGACTGTTTGAATATTCCAGGCAATATCGTGGCAC 415
 Db 328 AGAGGATATGCAAGTGTGTGTGATGCTGTGTCAACATCCCTGGCAATACCGGTGTAC 387
 QY 416 TTGTTTGTAGTGGTTCATGTTGGCTCATGACGGTGTCAATTTCTTGTATGTGGACGAGTG 475
 Db 388 CTCCTATGATGGATTCCACTGCACTGCACTGACGACACCACTGTCTGCA-GTGGACGAGTG 447
 QY 476 CTTGGGAACAAATGGCGGCTGCCAGGATACCTGTGTCAACGTCATGGGAGCTATGAGTG 535

Db 448 TGGCGAGGGCAACGGCGGCTGTGACGAGAGCTGTGTCAACATGATGGGAGCTATGAGTG 507
Qy 536 CTGCTCAAGAGGGGTTTTTCTGTAGTGACATCAGCACACCTGCAATTCACCGCTCGGA 595
Db 508 CCACCTCGCGGAAGGCTTCTTCTCAGCAGACACACAGCATACCTGTATCCAGCGGCCAG 567
Qy 596 AGAGGCGCTGAGCTGATGATTAAGGATCAGCGCTGTAGTCACATCTGCAAGAGAGGCCCC 655
Db 568 AGAAGGAATGAAATGATGAACAAGAACCAACGCGCTGTGCCACATTTGGCGGAGACACC 627
Qy 656 AAGGGCAGCGCTGCTGTGAGTGCAGGCGCTGTTTGTAGCTGGCCAAAGAACACAGAGAGA 715
Db 628 CAAGGGGGGTATTGCTGTGTAATGCCGTCTCTGCTTTGAGCTTACCAGAAACCAACGGGA 597
Qy 716 CTGCATCTTTGACCTGTAAACCATGGGAACGGTGGTGCAGCACTCTCTGTGACGATACAGC 775
Db 688 CTGTAAATTTGACATGCAACTATGTAAGCGCGCTGCCAGCACACCTGTGTGACACAGA 747
Qy 776 CGATGGCCAGAGTGCAGCTGCCATCGACGTACAGATGCCACACAGATGGGAGGAGCTG 835
Db 748 GCAGGGTCCCGGTGGGCTGCCATATCAAGTTTGTCTCATACCGACGGGAAGACATG 807
Qy 836 CTTTGAGCGAGAGGACACTGTCTGTGAGGTGACAGAGAGCAACACACATCACTGTGTGA 895
Db 808 CATCGGG-----GAAAGCGGCTAGAGCAGCACATCCCCACT-----844
Qy 896 TGGGGATAAACGGGTGAAACGGCGGCTGTCTATGGAAACGTGTGCTGTCAACAATGAGG 955
Db 845 -----CAAGCGGTTTCTAATGAGACCTGTGCTGTCAACAACGGGG 885
Qy 956 CTGTGACCGCAGCTGTAAAGATACTTGCAGAGTGTCCACTGCACTGCACTGTCTTGGATT 1015
Db 886 CTGTGACAGTAAGTGCCATGATCAGCGACTGTGTCCACTGCACCTGCCCTGTGGGCT 945
Qy 1016 CACTCTCCAGTTGATGGGAAGACATGTAAGATATTGATGAGTGCAGACCGCAATG 1075
Db 946 CATGCTGCACCCAGACNGGAAGAGCTGCCAAGATATAGATGAGTGGCGCTTAAACACGG 1205
Qy 1076 AGGTTGTGATCATTTCTGCAAAACATCGTGGGAGCTTTTGACTGGCGGTGCAAGAAAG 1135
Db 1006 GGGCTGTGACCATATTTGCCGCAACACAGTGGGCAGCTTCGATGCAGTTGCAAGAAAG 1065
Qy 1136 ATTTAAATTTAACAGATCAGAGAGTCTTCCAGATGTGGAAGAGTGTCTTTGATG 1195
Db 1066 CTATAAGCTTCTCATCAATGAGAGGAACCTGCCAGGATATAGACGAGCGTTCTTGTATG 1125
Qy 1196 GACTGTGACACAGCTGCATCAACCACTGACCACTTGTCTGTGCTTGCACACGAGG 1255
Db 1126 AACTGTGACCAATATGTGTCAACACCAAGAGGCTTCCAGTGTCTTGGCATCTGTG 1185
Qy 1256 GTACACCTGTATGGCTTCAACCACTGTGGAGACACCAATGAGTGCAGCATCAACAACGG 1315
Db 1186 CTACCTTGTGATGGTATCACCACTGTGGGATGTGGAAGATGCAAGTGCAGCATCAACGGGG 1245
Qy 1316 AGGCTGTACAGAGTGTGTGAAACACAGTGGGAGAGTGAATGTCAGTGCACACCTGTG 1375
Db 1246 AGGTTGGCGCTTGGCTGTGATCAACACTCTCTGGCAGCTACCACTGTACCTGCCAGCAGG 1305
Qy 1376 ---GTACAGCTCACTGGAATAAAGACTGTGTGGAAGTGAAGGGCTCTGCGCCAC 1432
Db 1306 CAGGCTGGCTGCACTGGAATGGCAAGATTTGCACAGAGCCACTGAAGTGTGAGGGCAG 1365
Qy 1433 AAGTGTGTACCCCGTGTGCTCTGCACTGCGGTGAAGAGTGTGGAGGAGACGGGTGCTT 1492
Db 1366 TCCTGGGCGCTCGAAGACCATGTCTCAGTGCACCGGCTGTGGCAGAGGACACCTGTGC 1425
Qy 1493 CCTCAGATGTACTCTGGCATTCACCTCTCTTCAGA-----1528
Db 1426 CCTCACCTGTCCCTCCAGGGGCCGATTTTGGCCAGAGCTGTGAGAAATGGCTTCAAGGTGAG 1485
Qy 1529 -----1528
Db 1486 CTGTGGGACCCCAAGCCAGGCTGTCTCCAGCCCGAGCTGGCCAAATGGGAACAGCAC 1545

Qy 1523 -----TGTCAACCAACATCAGGACAAAGTGTAACTTTAA 1561
Db 1546 CAACCTCCAAACCACTCCATGAGGCTGCACTGTCTCCATTAAACAACGGGCTCTCTCAA 1605
Qy 1562 GCTAAATGAAAGGCAAGTGTAGTTTGA-----AAATGCTGAGCTGTTTCCGAGGGTCT 1615
Db 1606 GATCAAGGATGCCAAATGCCCTTTCACCTCGCAACAAAGGCAAAACAGAGGAGGCTGG 1665
Qy 1616 GCGACCAACACTACACAGAAACACAGCTCAGTAAAGAGAGCTTCCCTACGTAAACCT 1675
Db 1666 CAGAACTCACAGGGCAGGTGTGCCCCCTGCTCTGATGCCAGGTCACTTCTCCACCT 1725
Qy 1676 TACATGCACTCTGCGCAAGCAAGTCCACAGAGCCCTTGGCCGACCAAGCAACCTTAAGGA 1735
Db 1726 TAAAGTCACTCTCTCGAAAGGGCAAGGGCCGACGGGCGGACCCCTCCAGGCAAGA 1785
Qy 1736 AATGTTTA---TCACTGTTGAGTTTGAGCTTGAACCTAAACAAAGGAGGTGACAGCTTC 1792
Db 1786 GGTCAAAAGGCTCACCTTGGAACTGGAGGCAAGGTTCAGAGCCGAAGAAACAACAGCCAG 1845
Qy 1793 TTGTGACCTGAGCTGCATCGTAAAGCGAAACCGAAGAGCGGTCCGTAAAGCCATCCGCAC 1852
Db 1846 CTG:GGGCTGCCCTGCCCTCCGACAGCAATGGAACGGCGGTGAAGAGATCCCTGAAGAT 1905
Qy 1853 GCTCAAAAGGGCGTCCACAGGGAGCAGTTTCACTCCAGCTCTCAGGCATGAACCTGA 1912
Db 1906 GCTCAGAAAGTCCATCAACCAAGGACCGCTCTGCTGCGGCTGGCAGGCTTGATTATGA 1965
Qy 1913 CTGTGCTAAAGGCTCCCAAGACATCTGAAGCCGAGCA-----GAGTCTCTGG 1963
Db 1966 GCTGCCCAACAGCCGGGCTGTGAGCCGGGAGGAGAGAGCGATGGAGTCTCTGTAG 2025
Qy 1964 AGTGGCCAGGGCTCATGCAGAAACCAATGTGTGAGTTCAGTTCGAGGGCTGGGACCTATTATGA 2023
Db 2026 GCCCGGCGAGCACCGTGTGGACCAAGTGTGAGCTGCCCGAGGGAACGTTATACCA 2085
Qy 2024 TGGAGCAGAGAACGCTGCATTTTATGTCCAAATGGAACTTCGAAATGAGGAAGACA 2083
Db 2086 CGGCCACAGCGAGCAGTGTGCCATGCCAGCGGCGACCTTCCAGGAGAGAGAAGGCA 2145
Qy 2084 ATGACTTGTGACCATGCCCAAGACAGGAAATCTGGGSCCTGAAAGACCCAGAGG 2143
Db 2146 SCTCTCTGCCACCTTGGCTTGGAG-----TGATGCCACGGGCTCTTGGAGC 2196
Qy 2144 TGGAAATGTCTGAATGTGAGGTCTGTCAACCTGGTGAATATTCGCAGAGTGGCTT 2203
Db 2197 CACCAAGCTCACCACTGTGCAGGTCACTGACCTGGCCACACTCTGTAGATGGGT 2256
Qy 2204 TGCACCTTGGCAGCTGTGTGCCCTGGGACGTTCCAGCTTGAAGCTGTGCAACTTCTGTG 2263
Db 2257 CAAGCCCTGTCAAGCATGCCACGCTGGCACCTACCACTTGAAGCAGGACGGACCTATG 2316
Qy 2264 CTTCCTCTGTGGAGAGGAGCTTGGCCACCAACATCAGGGAGTACTTCTTTCAGAGCTG 2323
Db 2317 CTTCCTCTGTG:GGGSCCTTCAACCAAGCATGAAGGGGCGCACTTCTTCCAGAGCTG 2376
Qy 2324 TGAAACCAAGATTCAATGTTCACCTGGACATTTTACAAACACCAACCTCACCGATGTAT 2383
Db 2377 TGACACCAAGTCCAGTGTGCCAGGGCAGTACTTACAAACACCAAGTCCACCTGTAT 2436
Qy 2394 TCGTTGCCAGTGGGAACATACCAAGCTGAATTTGGAAAAATAATTTGTGTTCTTGGCC 2443
Db 2437 TCGCTGTGCATGGGCTCCTATCAGCCGACTTCCGTCAAGAACTTCTCAGCGGCTGTC 2496
Qy 2444 AGGAATACTACGACTGACTTTGATGGCTTCACAAACATAACCCAGTGTAAAAACAGAG 2503
Db 2497 AGGAACAACAGCAGACTTTGATGGCTTACCAAGTGTGGCCCAATGCAAGAAATCTGCA 2556
Qy 2504 ATGTGGAGGGAGCTGGAGATTTTCACTGGGTACATTTGAAATGCCCAAACTACCGAGGCA 2563
Db 2557 GTGTGGTGGGAGCTGGGTGAGTTCACTGGCTATATTGAGTTCACCACTACCGGGCA 2616


```
QY 2564 TTACCCAGCCAAACACCGAGTGTAGTGGACCATCAACCCACCCGCCAAGCGCGGATCCT 2623
Db 2617 CTACCCAGCTGGTGTGGAGTGCATCTGGAAATCAACCCGCCAAGCGGAAGATCCT 2676
QY 2624 GATCGTGGTCCCTGAGATCTTCTGCCCATAGAGGACGACTGTGGGACTATCTGGTGAT 2683
Db 2677 TATCGTGGTACCAGAGATCTTCTGCCATCTGAGGATGAGTGTGGGACGTCCTCGTCAT 2735
QY 2684 GCGAAACCTCTTCAZTCCAAATCTGTGACAAATATGAAACCTGCGACACCTACGAAAG 2743
Db 2737 GAGAAAGAACTCATCCCATCTCCATTAACCACTTATGAGACCTGCGACACCTACGAGG 2796
QY 2744 CCCATCGCCTTCACTCCAGTCAAGAGCTGTGGATTTCAGTTCCAAATGAAAG 2803
Db 2797 TCCATTTGCTTCACTGCCGCTCCAGGAAGCTCTGTGATCAACTTCAAGAAAGCGAGC 2856
QY 2804 GAACAGCGCTAGAGGTTCCAGGTCCCATACGTGACATATATGAGGACTACCGAAT 2863
Db 2857 CAACAGCGCGCTGGCTTCCAGATTCCCTATGTACCTATGATGAGGACTATGAGCAGCT 2916
QY 2864 CATTGAAGACATATGTCAGATGGCAGGCTCTATGCACTGAGAACCATCAGGAATACT 2923
Db 2917 GGTAGAAGACATTTGTGGAGATGCGCGCTCTATGCTCTGAAAGACACGAGGATTT 2976
QY 2924 TAAGGATAAGAACTTATCAAGGCTGTGTTGATGCTCTGCGCCATCCCGAGACTATTT 2983
Db 2977 AAGGACAAAGAGCTCATCAAGGCTTCTTTGAGTGTCTAGCCACCCCGAGAACTACT 3036
QY 2984 CAAGTACACAGCCCGAGAGTCCCGAGAGATGTTTCCAAGATCGTTTCATCGGATGTCG 3043
Db 3037 CAAGTACACAG--AGAAACAAAGAGAGATGCTGCCAATAATCTTCATCAAGCTGCTCG 3093
QY 3044 TCCAAAGTCCAGGTTTGTGACCTTACAAT 3078
Db 3094 CTCCAAAGTTCCAGCTTCTGAGGCGCTACAAT 3128

RESULT 15
US-09-898-570-9
; Sequence 9, Application US/09898570
; Patent No. US20020123612A1
; GENERAL INFORMATION:
; APPLICANT: GERLACH, VALERIE L.
; APPLICANT: ELLERMAN, KAREN
; APPLICANT: MACDOUGALL, JOHN R.
; APPLICANT: SMITHSON, GLENDA
; TITLE OF INVENTION: NOVEL HUMAN PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND
; FILE OF INVENTION: METHODS OF USING THE SAME
; FILE REFERENCE: 15966-776CIP
; CURRENT APPLICATION NUMBER: US/09/898,570
; CURRENT FILING DATE: 2001-07-03
; PRIOR APPLICATION NUMBER: 60/198,293
; PRIOR FILING DATE: 2000-04-19
; PRIOR APPLICATION NUMBER: 60/198,645
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: 60/210,809
; PRIOR FILING DATE: 2000-06-09
; PRIOR APPLICATION NUMBER: 60/199,476
; PRIOR FILING DATE: 2000-04-26
; PRIOR APPLICATION NUMBER: 60/200,025
; PRIOR FILING DATE: 2000-04-26
; PRIOR APPLICATION NUMBER: 60/224,610
; PRIOR FILING DATE: 2000-08-11
; PRIOR APPLICATION NUMBER: 60/200,024
; PRIOR FILING DATE: 2000-04-26
; PRIOR APPLICATION NUMBER: 60/199,880
; PRIOR FILING DATE: 2000-04-26
; PRIOR APPLICATION NUMBER: 60/218,591
; PRIOR FILING DATE: 2000-07-17
; PRIOR APPLICATION NUMBER: 60/271,914
; PRIOR FILING DATE: 2001-02-27
; PRIOR APPLICATION NUMBER: 60/215,855
; PRIOR FILING DATE: 2000-07-03
```

```
; PRIOR APPLICATION NUMBER: 09/839,446
; PRIOR FILING DATE: 2001-04-19
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 9
; LENGTH: 4440
; TYPE: DNA
; ORGANISM: Unknown Organism
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: POLYX
; OTHER INFORMATION: 297812_8.0.7C4
US-09-898-570-9

Query Match 25.23; Score 340.8; DB 10; Length 4440;
Best Local Similarity 6.11; Pred. No. 9.7e-277; Indels 192; Gaps 9;
Matches 1791; Conservative 0; Mismatches 947;

QY 176 CCCCGCGGTGGGGCGGTGCGCGGGCGCGGAGGAGTGTAGATGAGTGTGCCCAAGG 235
Db 120 CACCCCGCGCGCGCGCGAGTACAGCAAGCGCG-GCAAGATGTGATGAGTGTGGAGG 178
QY 236 GCTAGATGACTGCCATGCCAGCGCCCTGTGTGAGAACACACCCACCTCTTACAAAGTCTC 295
Db 179 GACTGACAACTGCCACATCGATGCTATCTGCCAGAACACCCCGAGGTCTACAAAGTGCAT 238
QY 296 CTGCAAGCCTGGCTACCAAGGGGAGGAGGAGGAGTGTGAGGACATCATGATGATGTA 355
Db 239 CTGCAAGTCTGGCTACACAGGGGAGCGCAACACTGCCAAGAGAGTGTGATGAGTGCAGG 298
QY 356 TGAGTCAATGAGGAGGTGTGTCCATGACTGTTTAAATATTCAGGCAATATGTTTGCAC 415
Db 299 AGAGGATTAATCGAGTGTGTGCAATGACTGTGTCAACATCCCTGGCAATACCGGTGTAC 358
QY 416 TTGTTTATGCTTCAATGTTGGCTCATGAGCGTCAATTTGCTTCAATGAGTGTGAGTGTG 475
Db 359 CTGCTATGATGGATTCCACCTGGCACATGACGAGCAACAATGTCTGATGTGACGAGTG 418
QY 476 CTTGAGAACAAATGCGGCTGCCAGCATACCTGTGTCAACGTCACTGGGAGCTATGAGTG 535
Db 419 TGCCGAGGGCAACGCGGCTGTACAGAGAGCTGTGTCAATGATGGCGAGCTATGAGTG 478
QY 536 CTGTGCAAGGAGGGGTTTTTCTTGAGTGAATACAGCACACCTGCAATCACCCCTCGGA 595
Db 479 CCATCTGCGGGAAGGCTTCTTCTCAGGACAAACAGCATACCTGTATCCAGCGCCAGA 538
QY 596 AGAGGCTCTGAGTGTGATGAATTAAGATCAAGGCTGTAGTCAATCTGCAAGGAGGCGCC 655
Db 539 AGAAGGAATGAATTTGCATGAAACAAGAACCCAGGCTGTGCCACATTTGCCGGGAGACCC 598
QY 656 AAGGGGACGCTCGCCTGTGAGTGCAGGCTGGTTTTGAGCTGCCCAAGAACACAGAGAGA 715
Db 599 CAAGGSGGATATGCTTGTGAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 658
QY 716 CTGCATCTTGACCTGTAAACCATGGGAACGGTGGGTGCCAGACCTCTGTGCAAGATCAGC 775
Db 659 CTGTAAATTGACATGCAACTATGTTAAACGGCGCTGCCAGCACACGCTGTGATGACAG- 717
QY 776 CGATGGCCCAAGTGCAGCTGCCATCCACATCAAGATGCACACAGATGGAGGAGCTG 835
Db 718 ----- 717
QY 836 CTTTGAGCGAGAGGACACTGTCTCTGGAGGTGACAGAGAGCAACACACATCAGTGGTGA 895
Db 718 ----- 717
QY 896 TGGGATAAAGCGGTGAAGAGCGGCTGCTCATGGAACGTTGTCTGTCAACAAATGAGG 955
Db 718 ----- 717
QY 956 CTGTGACGCGACCTGTGAAGATACCTTCAGAGAGTGTTCAGTGTGAGTGTGCTGTGATTT 1015
Db 747 CTGTGACAGTAAAGTSCCATGATGAGGAGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1015
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 22, 2003, 17:27:46 ; Search time 238 Seconds

(without alignments)
6930.457 Million cell updates/sec

Title: US-09-747-371-1

Perfect score: 3737

Sequence: 1 ggcgccgcgcacaccc.....ctaagcactctggagacat 3737

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents NA.*
1: /cgn2_6/ptodata/2/ina/5A COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PTUS COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	257	6.9	291	4	US-09-389-681-246
C 2	257	6.9	291	4	US-09-620-4058-246
C 3	257	6.9	291	4	US-09-339-338-246
C 4	257	6.9	291	4	US-09-433-8269-246
C 5	257	6.9	291	4	US-09-604-287A-246
6	213.4	5.7	241	4	US-09-439-313-412
7	213.4	5.7	241	4	US-09-352-616A-412
8	83	2.2	3373	2	US-08-897-443-2
9	78.6	2.1	1531	2	US-08-833-963C-1
10	78.6	2.1	2018	3	US-08-380-514-2
11	76.6	2.0	1254	4	US-09-312-283C-67
12	75	2.0	1260	3	US-09-188-930-67
13	71.8	1.9	2550	2	US-08-884-072-2
14	71.8	1.9	2550	4	US-09-312-168-2
15	70.2	1.9	2461	1	US-08-282-141-1
16	63.6	1.7	1018	3	US-09-188-930-259
17	63.6	1.7	1018	4	US-09-312-283C-259
18	61.4	1.6	3546	3	US-08-872-757-3
19	61.4	1.6	3546	4	US-09-850-048A-3
20	59.4	1.6	2512	4	US-09-248-757-1
21	58.8	1.6	3284	6	5258288-3
22	58.8	1.6	3290	1	US-07-985-691-1
23	58.8	1.6	3290	1	US-08-436-804-1
24	58.8	1.6	3290	1	US-08-267-387-1
25	58.8	1.6	6811	3	US-08-651-472-67
C 26	58.8	1.6	6811	3	US-08-358-928-67
27	56.4	1.5	2397	6	5258288-2

28	55.9	1.5	2288	4	US-09-620-312D-24
29	54	1.4	5089	6	5177-97-31
C 30	53.6	1.4	220	4	US-09-404-879A-340
31	53.4	1.4	3690	3	US-08-991-408-3
32	53.4	1.4	3690	4	US-09-432-473-3
33	53.4	1.4	3919	2	US-08-866-650-4
34	53.4	1.4	3919	2	US-09-021-287-4
35	53.4	1.4	3919	3	US-09-240-473-4
36	53.4	1.4	5145	3	US-08-391-408-1
37	53.4	1.4	5145	4	US-09-432-473-1
38	52.6	1.4	4661	4	US-09-285-385C-3
39	52.6	1.4	477	2	US-08-866-650-2
40	52.6	1.4	477	2	US-09-021-287-2
41	52.6	1.4	4771	3	US-09-240-473-2
42	52.2	1.4	502	4	US-09-285-385C-1
C 43	52	1.4	2056	3	US-09-334-602-2
44	51.4	1.4	1203	3	US-09-586-01C-1
45	50.6	1.4	3759	3	US-08-479-722B-3

ALIGNMENTS

RESULT 1

US-09-389-681-246/c

; Sequence 246, Application US/09389681A

; Patent No. 6519237

; GENERAL INFORMATION:

; APPLICANT: Yuqiu, Jiang

; APPLICANT: Dillon, Davin C.

; APPLICANT: Mitcham, Jennifer L.

; APPLICANT: Xu, Jiangchun

; TITLE OF INVENTION: COMPOSITIONS FOR THE TREATMENT AND

; TITLE OF INVENTION: DIAGNOSIS OF BREAST CANCER AND METHODS FOR THEIR USE

; FILE REFERENCE: 210121.470C3

; CURRENT APPLICATION NUMBER: US/09/389,681A

; CURRENT FILING DATE: 1999-09-02

; NUMBER OF SEQ ID NOS: 463

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 246

; LENGTH: 291

; TYPE: DNA

; ORGANISM: Homo sapiens

; NAME/KEY: misc feature

; LOCATION: 11..(291)

; OTHER INFORMATION: n=A,T,C or G

US-09-389-681-246

Query Match: 6.9%; Score 257; DB 4; Length 291;

Best Local Similarity 97.5%; Pred. No. 2.2e-57;

Matches 268; Conservative 6; Indels 1; Gaps 1;

QY 2506 GCCTGGGACCTATTATGATGGACGACGAGAACCG-TGCATTTTATGTCANATGGAACCT 2064

DB 275 GGCTGGGACCTATTATNATNAGCAGAGAACCTTTCATTTTATGTCANATGGAACCT 216

QY 2065 TCCAAATGAGGAGGACAAATGACCTGTGCAACCATGCCACAGCAGGAAATTCGGGG 2124

DB 215 TCCAAATGAGGAGGAGGACAAATGACCTGTGCAACCATGCCACAGCAGGAAATTCGGGG 156

QY 2.25 CCCTGAAGACCCAGAGCTTGGAAATATGTCGAAATGTGGAGGTCTGTGTCAACCTGGTG 2184

DB 155 CCCTGAAGACCCAGAGCTTGGAAATATGTCGAAATGTGGAGGTCTGTGTCAACCTGGTG 96

QY 2185 AATATTCGAGATGGCTTGGACCTTGCACGCTCTGTGCCCTGGGACGCTTCCAGCCTG 2244

DB 95 AATATTCGAGAGNGCTTGCACCTTGCACGCTCTGTGCCCTGGGACGCTTCCAGCCTG 36

QY 2245 AAGCTGGTGAACCTTCTGCTGCTTCCCTGTGGAGGA 2279

DB 35 AAGCTGGTGAACCTTCTGCTGCTTCCCTGTGGAGGA 1

Best Local Similarity 98.7%; Pred. No. 4.5e-46;
Matches 225; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 2350 GACATTTCTACAAACACCACTACACCGATGATTTCGTTGCCAGTGGGAACATACCAGC 2409
DB 15 GACATTTCTACAAACACCACTACACCGATGATTTCGTTGCCAGTGGGAACATACCAGC 73
QY 2410 CTGAATTTGGAAAAAATAATTGTGTTTCTGCCAGGAAATACACGACTGTTGATG 2469
DB 74 CTGAATTTGGAAAAAATAATTGTGTTTCTGCCAGGAAATACACGACTGTTGATG 133
QY 2470 GCTCCACAAACATACCCAGTGTAAACACAGAGNATGAGGGGAGCTGGAGATTTC 2529
DB 134 GCTCCACAAACATACCCAGTGTAAACACAGAGNATGAGGGGAGCTGGAGATTTC 193
QY 2530 CTGGGTACATTCGAATCCCAACTACCCAGCAATTCACCCAGCCAACA 2577
DB 194 CTGGGTACATTCGAATCCCAACTACCCAGCAATTCACCCAGCCAACA 241

RESULT 8

US-08-897-443-2
; Sequence 2, Application US/08897443
; Patent No. 5981263
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Lal, Preeti
; APPLICANT: Corley, Neil C.
; APPLICANT: Shah, Purvi
; APPLICANT: Kaser, Mathew
; TITLE OF INVENTION: HUMAN XATRILIN-3
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/897,443
; FILING DATE: Filed Herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0348 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3373 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: UTRNOT02
; CLONE: 681719
US-08-897-443-2

Query Match 2.2%; Score 83; DB 2; Length 3373;

Best Local Similarity 50.7%; Pred. No. 1.6e-11;
Matches 253; Conservative 0; Mismatches 240; Indels 6; Gaps 2;

QY 925 TCATGGAACGTTGCTGTCAACAATGGAGGCTGTGACCGCACCTGTGAAGGACTACTTCA 984

DB 890 TCAGGATCTGTGCCATGGAGGACCAAACTGTGACGAGCTGTGTGAATGTGCCG 949
QY 985 CAGGTGTCCACCTGCAGTTGTCTCTGTGGATTCACTCTCCAGTTGGTGGGAAGACATGTA 1044
DB 950 GTCCTTTCTGTCGCGAGTGCTACAGTGGCTACGGCTGGCTGAGGATGGGAAGAGGTG 1009
QY 1045 AAGATATTGATGAGTGCAGACCCGCAATGGAGGTTGTGATCATTTCTGCAAAAACATCG 1104
DB 1010 TGGCTGTGAGCTACTGTSCCTCAGAAAACACCGATGTGAACATGATGTGTAAATGCTG 1069
QY 1105 TGGGAGATTTTGAATGCGGCTGGAAGAAAGGATTTAAATTATTAACAGATGAGAGTCTT 1164
DB 1070 ATGGCTCTCTACTTTTGGCAGTGCCATGAAGGATTTGTCTTAACCCAGATAAAAAACGT 1129
QY 1165 GCCAAGATGTGATGAGTG---CTCTTTGGATAGGACTGTGACCAAGCTGCAATCAACC 1221
DB 1130 GCCAAGATAGACTACTGTGCTTCATCTAATCAGGATGTGACGAGAGTGTGTTAACA 1199
QY 1222 ACCCTGGCACATTTGCTTGTCTTGCACCGAGGATACACCTGTAT---GGTTTCAACC 1278
DB 1190 CAGATGATTCCTATTCCTGCCACTGCTGAAAGGCTTTACCTGAATCCAGATAAGAAA 1249
QY 1279 ACTGTGGAGACACCAATGAGTGCAGCATCAACACGAGGCTGTGACGAGTCTGTGTA 1338
DB 1250 CTTGAGAGGATCAACTACTGTGCACTGAACAAACCGGCTGTGAGCATGAGTSCGTCA 1389
QY 1339 ACACAGTGGCAGCTATGATGCCAGTGCACCTGGGTACAAAGCTCCACTGGGAATAAA 1398
DB 1310 ACATGGAGAGAGCTACTACTGCGCTGCCACCGTGGCTACTCTGACCCCAATGGCA 1368
QY 1399 AAGACTGTGTGGAAGTAA 1417
DB 1370 AAACCTGCAGCGAGTGA 1388

RESULT 9

US-08-933-963C-1
; Sequence 1, Application US/08833963C
; Patent No. 5916769
; GENERAL INFORMATION:
; APPLICANT: Olsen, et al.
; TITLE OF INVENTION: Extracellular/Epidermal Growth Factor
; TITLE OF INVENTION: HCABAS8X
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Ave
; CITY: Rockville
; STATE: MD
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/933,963C
; FILING DATE: 11-APR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WC PCT/US96/05033
; FILING DATE: 13-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: P258
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301-309-8534
; TELEFAX: 301-309-8439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:

LENGTH: 1531 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 125..1453
US-08-833-963C-1

Query Match 2.1%; Score 78.6; DB 2; Length 1531;
Best Local Similarity 54.6%; Pred. No. 1.5e-10;
Matches 206; Conservative 0; Mismatches 159; Indels 12; Gaps 2;
QY 210 GAGGATGTAGATGAGTGTGCCCAAGGGCTAGATGACTGCCATGCCAGCCCTCTGTGTGAG 269
DB 488 GTGGATGTGGACGAGTGTGCCAGAGCCCTGCACGACTGTGCCCCAGCCAGGACTGCCAT 547
QY 270 AACACACCCACCTCTTACAAAGTGTCTGCAAGCCTGCTACCAAGGGGAAAGGAGGAGCAG 329
DB 548 AACTTGGCTGCTCTATCATGAGTGCACCTGCCCTGATGCTTACCGCAAGATCGGGCCGAG 607
QY 330 TGTGAGGACATCGAATGTGGAATGAGCTCAATGAGGCTGTGTCATGACTGTGTTG 389
DB 608 TGTGAGGACATGACGAGTG-----CCGCTACCCCTACTGCCAGCAGCCGCTGGTG 658
QY 390 AATATTCCAGGCAATATCGTTGCACTTGTGTTTGTGATGGCTTCATGTTGGCTCATGACGGT 449
DB 659 AACCTGCTGGCTCTCCGCTGCAGTGCAGCGCGGCTTCCAGCTGGGCGCTAACAC 718
QY 450 CATAATTGCTTGTGAGTGTGAGTGTGCTGAGGAAATGAGGCTGCGGAGTGCACGACTACCTGT 509
DB 719 CGCTCTCTGTTGATGTGAACAGTG---TGACATGGGGGCGCCCATGCGAGCAGCGGTGC 775
QY 510 GTCACAGCTCATGGGAGCTATGAGTGTGCTGCAAGGAGGGGTTTTCCTGAGTGACAAAT 569
DB 776 TTCAACTCTATGGAGCTTCTGTGCTGCCACAGGCTATGAGCTGCATCGGGAT 835
QY 570 CAGCACACCTGCATTCA 586
DB 836 GCGCTTCTCTGCAGTGA 852

RESULT: 10
US-08-980-514-2
Sequence 2, Application US/089805:14
Patent No. 6004753
GENERAL INFORMATION:
APPLICANT: Yue, Henry
APPLICANT: Guegler, Karl J.
APPLICANT: Shah, Purvi
TITLE OF INVENTION: HUMAN Si-5-ECMP-LIKE PROT
TITLE OF INVENTION: EIN
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSES: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Dr.
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/980,514
FILING DATE: Filed Herewith
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-C436 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2018 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: BRSTNOT13
CLONE: 2786449
US-08-980-514-2
Query Match 2.1%; Score 78.6; DB 3; Length 2018;
Best Local Similarity 54.6%; Pred. No. 1.7e-10;
Matches 206; Conservative 0; Mismatches 159; Indels 12; Gaps 2;
QY 210 GAGGATGTAGATGAGTGTGCCCAAGGGCTAGATGACTGCCATGCCAGCCCTCTGTGTGAG 269
DB 569 GTGGATGTGGACGAGTGTGCCAGAGCCCTGCACGACTGTGCCCCAGCCAGGACTGCCAT 628
QY 270 AACACACCCACCTCTTACAAAGTGTCTTGCAGCCTGCTACCAAGGGGAAAGGAGGAGCAG 329
DB 629 AACTTGGCTGCTCTATCATGAGTGCACCTGCCCTGATGCTTACCGCAAGATCGGGCCGAG 688
QY 330 TGTGAGGACATCGAATGTGGAATGAGCTCAATGAGGCTGTGTCATGACTGTGTTG 389
DB 689 TGTGAGGACATGACGAGTG-----CCGCTACCCCTACTGCCAGCAGCCGCTGGTG 739
QY 390 AATATTCCAGGCAATATCGTTGCACTTGTGTTTGTGATGGCTTCATGTTGGCTCATGACGGT 449
DB 740 AACCTGCTGGCTCTCCGCTGCAGTGCAGCGGGCTTCCAGCTGGGCGCTAACAC 799
QY 450 CATAATTGCTTGTGATGTGAGCAGTGTGCTGAGGAAATGAGGCTGCGGAGTGCACGACTACCTGT 509
DB 800 CGCTCTCTGTTGATGTGAACAGTG---TGACATGGGGGCGCCCATGCGAGCAGCGCTGC 856
QY 510 GTCACAGCTCATGGGAGCTATGAGTGTGCTGCAAGGAGGGGTTTTCCTGAGTGACAAAT 569
DB 857 TTCAACTCTATGGAGCTTCTGTGCTGCCACAGGCTATGAGCTGCATCGGGAT 916
QY 570 CAGCACACCTGCATTCA 586
DB 517 GCGCTTCTCTGCAGTGA 933

RESULT: 11
US-09-312-283C-67
Sequence 67, Application US/09312283C
Patent No. 6573095
GENERAL INFORMATION:
APPLICANT: Watson, James D.
APPLICANT: Strachan, Lorna
APPLICANT: Sleeman, Matthew
APPLICANT: Orust, Rere
APPLICANT: Marison, James G.
APPLICANT: Kurbie, Krishanand D.
TITLE OF INVENTION: Compositions Isolated from Skin Cells
TITLE OF INVENTION: and Methods for Their Use
FILE REFERENCE: 11005.1011c2
CURRENT APPLICATION NUMBER: US/09/312,283C
CURRENT FILING DATE: 1999-05-14
NUMBER OF SEQ ID NOS: 425
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 67
LENGTH: 1254
TYPE: DNA
ORGANISM: Mouse
US-09-312-283C-67

Query Match 2.0%; Score 76.6; DB 4; Length 1254;
Best Local Similarity 53.0%; Pred. No. 4.5e-10;
Matches 188; Conservative 0; Mismatches 164; Indels 3; Gaps 1;
QY 1068 CGCAATGAGGTTGTGATCATTTCTGCAAAACATCGTGGCAGCTTTTGTGATCGGGCTGC 1127
DB 180 CGCTATGTTTACTGCCAGCAGCTCTGTGCGAATGTTCTCGATCTATCTCTGTACCTGT 239
QY 1128 AAGAAAGATTAAATTAATTAACAGATGAGAAGTCTTGCCAAAGATGTGGATGTCTCT 1187
DB 240 AACCTGGCTTCAACCTCAACGATGATGCAAGTCTTCCCAAGATGGAACGAGTGTGA 299
QY 1188 TTGNATGAGCTGTGACCAAGCTGTGATCAACACCTCGCACATTTGCTTGTGCTTGC 1247
DB 300 ACTGAGAACCCCTGTGTCAGACCTGCGTCAACACCTATGGTCTTTCATCTGCCCTGT 359
QY 1248 AACCGAGGTACACCT---GTATGGCTTCCACCACTGTGGAGACACCAATGAGTGCAGC 1324
DB 360 GACCCAGATATGAATCGAGGAGATGGCATTCTACTGCATGATATGATGAGTGCAGC 419
QY 1305 ATCAACACGAGGCTGTGACAGGTCTGTGTGAACACAGTGGCAGCTATGAATGCCAG 1364
DB 420 TTCTCCGAGTTCCTGTCAACATGAGTGTGTGAACACGCGGCTCATCTCTCTGCTCA 479
QY 1365 TGCCACCTGGGTACAACTCCACTGGATATAAAGAGCTGTGGAAGTGAAGG 1419
DB 480 TGCCCTCCAGCTACGTCTTGTGGAAGATAACCGAAGCTCCAGGATATCAATG 534
RESULT 12
US-09-188-930-67
; Sequence 67, Application US/09188930A
; Patent No. 6150502
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Carust, Rene
; APPLICANT: Murison, James Greg
; TITLE OF INVENTION: Compositions Isolated From Skin Cells
; TITLE OF INVENTION: and Methods For Their Use
; FILE REFERENCE: 11000.1011c1
; CURRENT APPLICATION NUMBER: US/09/188,930A
; CURRENT FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 348
; SOFTWARE: FastSeq For Windows Version 3.0
; SEQ ID NO 67
; LENGTH: 1260
; TYPE: DNA
; ORGANISM: Rat
US-09-188-930-67
Query Match 2.0%; Score 75; DB 3; Length 1260;
Best Local Similarity 51.8%; Pred. No. 1.2e-09;
Matches 184; Conservative 4; Mismatches 164; Indels 3; Gaps 1;
QY 1068 CGCAATGAGGTTGTGATCATTTCTGCAAAACATCGTGGCAGCTTTTGTGATCGGGCTGC 1127
DB 186 CGCTATGTTTACTGCCAGCAGCTCTGTGCGAATGTTCTCTGGAATCTATTTCTCTGTACGTGT 245
QY 1128 AAGAAGAGATTAAATTAATTAACAGATGAGAAGTCTTGCCAAAGATGTGGATGTCTCT 1187
DB 246 AACCTGGCTTCAACCTCAACGATGATGGAAGTCTTGCCAAAGATGTGAACGAGTGTGA 305
QY 1188 TTGGATAGACCTGTGACCAAGCTGCATCAACACCTCGCACATTTGCTTGTGCTTGC 1247
DB 306 ACTGAGAACCCCTGTGTTTCAGACCTGGGTCACACCTATGGTTCCTTCATCTGCCGCTGT 365
QY 1248 AACCGAGGTACACCT---GTATGGCTTCCACCACTGTGGAGACACCAATGAGTGCAGC 1304
DB 366 GACCCAGKATATGAATGAGGAAATGGCATTCACCTGCATGATATGGAATGAGTGCARC 425

QY 1305 ATCAACACGAGGCTGTGATCATTTCTGCAAAACATCGTGGCAGCTATCAATGCCAG 1364
DB 426 TTCTCCGAGTTCCTGTGCAACATGAGTGTGTGAACACGCGGCTCATCTCTGCTCA 485
QY 1365 TGCCACCTGGGTACAAAGCTCCACTGGAATAAAGAGCTGTGTGGAAGTGAAGG 1419
DB 486 TGCCCTCCAGGCTWCKTCTTGTGGAAGATAACCGAAGCTCCAGGATATCAATG 540
RESULT 13
US-08-984-072-2
; Sequence 2, Application US/08884072
; Patent No. 5872234
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Corley, Neil C.
; APPLICANT: Guegler, Karl J.
; TITLE OF INVENTION: HUMAN EXTRACELLULAR MATRIX PROTEINS
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/884,072
; FILING DATE: Herewith
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0333 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2550 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: CORNOT01
; CLONE: 45517
US-08-984-072-2

Query Match 1.9%; Score 71.8; DB 2; Length 2550;
Best Local Similarity 52.1%; Pred. No. 1.1e-08;
Matches 185; Conservative 0; Mismatches 167; Indels 3; Gaps 1;
QY 1068 CGCAATGAGGTTGTGATCATTTCTGCAAAACATCGTGGCAGCTTTTGTGATCGGGCTGC 1127
DB 898 CGCTATGTTTACTGCCAGCAGCTCTGTGCGAATGTTCTCTGGATCTCTATCTTGTACATGC 957
QY 1128 AAGAAGAGATTAAATTAATTAACAGATGAGAAGTCTTGCCAAAGATGTGGATGTCTCT 1187
DB 958 AACCTGGTGTACCTCAATGAGGATGGAAGCTCTTGCCAAAGATGTGAACGAGTGTGCC 1017
QY 1188 TTGGATAGACCTGTGACCAAGCTGCATCAACACCTCGCACATTTGCTTGTGCTTGC 1247
DB 1018 AACGAGAACCCCTGCTGCAAAACCTGCGTCAACACCTACGGCTCTTTCATCTGCGGCTGT 1077

QY 1248 AACCGAGGTACACCT---GTATGGCTTACCCACTGTGAGACACCAATGAGTGCAGC 1364
DB 1078 GACCCAGGATGTAACCTTGAGGAAGATGGCGTTCATTGCAATGATATGACGAGTGCAGC 1137
QY 1305 ATCAACACGAGGAGTGTGACGAGGTGTGTGAACACAGTGGGAGCTATGAATGCCAG 1364
DB 1138 TTCTCTGAGTTCTCTGCCACATGAGTGTGTAACACCGCCGGCACATATTTCTGCTCC 1197
QY 1365 TGCCACCTGGGTACAAAGTCCACTGGGAATAAAAGAACTGTGTGAAGTGAAG 1419
DB 1198 TGCCCTCCAGGTACATCTCTGCTGATGACACCGCAAGCTGCCAAGACATCAAGC 1252

RESULT 14

US-09-212-168-2
; Sequence 2, Application US/09212168
; Patent No. 6303765
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Corley, Neil C.
; APPLICANT: Guegler, Karl J.
; TITLE OF INVENTION: HUMAN EXTRACELLULAR MATRIX PROTEINS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Inocyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/212,168
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/884,072
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0333 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2550 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: CORNNOT01
; CLONE: 45517

US-09-212-168-2

Query Match 1.9%; Score 71.8; DB 4; Length 2550;
Best Local Similarity 52.1%; Pred. No. 1.1e-08;
Matches 185; Conservative 0; Mismatches 167; Indels 3; Gaps 1;
QY 1068 CGCAATGGAGGTGTGATCATTTCTCAAAAACATCGTGGGAGTTTGACTGCGGTCG 1127
DB 898 CGCTATGGTTACTGCCAGGAGCTCTGTGCGAATGTTCTCTGGATCTCTATTCTTGATACGC 957
QY 1128 AAGAAAGGATTAATTAATTAACAGATGAGAACTTTTGCAAGATGTGGATGAGTCTCT 1187
DB 958 AACCTGGTTTACCCTCAATGAGGATGGAAGTCTTGGCAAGATGTGAACGAGTGTGCC 1017

QY 1188 TTGGATAGACCTGTGACGACAGCTGCATCAACCCCTGGACACATTTGCTGTGCTTGC 1247
DB 1018 AACGAGAACCCCTGCTGCAAACTCGTCAACACCTACGGCTCTTTATCTGCGCGCTGT 1077
QY 1248 AACCGAGGTTACACCT---GTATGGCTTACCCACTGTGAGACACCAATGAGTGCAGC 1304
DB 1078 GACCCAGGATGTAACCTTGAGGAAGATGGCGTTCATTGCAATGATATGACGAGTGCAGC 1137
QY 1305 ATCAACACGAGGAGTGTGACGAGGTGTGTGAACACAGTGGGAGCTATGAATGCCAG 1364
DB 1138 TTCTCTGAGTTCTCTGCCACATGAGTGTGTGAACACCGCCGGCACATATTTCTGCTCC 1197
QY 1365 TGCCACCTGGGTACAAAGTCCACTGGGAATAAAAGAACTGTGTGAAGTGAAG 1419
DB 1198 TGCCCTCCAGGTACATCTCTGCTGATGACACCGCAAGCTGCCAAGACATCAAGC 1252

RESULT 15

US-06-282-141-1
; Sequence 1, Application US/08282141
; Patent No. 5538861
; GENERAL INFORMATION:
; APPLICANT: Schneider, Claudio
; APPLICANT: Varum, Brian
; APPLICANT: Avanzi, Giancarlo
; APPLICANT: Brancolini, Claudio
; APPLICANT: Manfioletti, Guido
; TITLE OF INVENTION: Stimulating Factor for the AXI Receptor
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Aagen Inc.
; STREET: 1840 Denaville Drive
; CITY: Thousand Oaks
; STATE: California
; COUNTRY: United States
; ZIP: 91320-1799
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.3, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/282,141
; FILING DATE:
; CLASSIFICATION: 435
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2461 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; US-08-282-141-1

Query Match 1.9%; Score 70.2; DB 1; Length 2461;
Best Local Similarity 54.1%; Pred. No. 2.9e-08;
Matches 166; Conservative 0; Mismatches 138; Indels 3; Gaps 1;
QY 946 ACAATGGAGCTGTGACCGCACCTGTAAAGATATCTTCGACAGGTGTCACCTGCAGTTGTC 1005
DB 622 AGAAGCGGGCTGCTCCAGATCTGCCACAAAGCGGGTAGCTTCCACTGTTCTCTGCC 681
QY 1006 CTGTTGATTCACCTCCAGTGTGGAGAACATGTAAAGATATGATGAGTGCAGCA 1065
DB 682 ACAGCGCTTCGAGCTCTCTCTGATGGCAGGACCTGCCAAGACATAGACGAGTGGCGAG 741
QY 1066 CCGCATGAGGTTGTGATCATTTCTGCAAAACATCGTGGGCACTTTGACTGCGGCT 1125
DB 742 ACTCGGAGGCTCGGGAGGCGCGCTGCAAGAACCTGCCCGGCTCTACTCTGCTCT 801
QY 1126 GCAAGAAAGGATTAATTAATTAACAGATGAGAAGTCTTGCCAAAGATGTGGATGAGTCT 1185
DB 932 GTACAGGAGGCTTTGCTACAGCTCCAGAGAGAGCTTGGCAGATGTGGACAGATG-- 859

QY 1186 CTTTGGATAGGACCTGTGACACAGCTGCATCAACACCCCTGGCACATTTGCTGTGCTT 1245
Db 860 -TCTGACGGCGCTGTGAGCAGGTCTGCTGAATCTCCCAAGGAGCTACACCTGCCACT 918
QY 1246 GCAACCG 1252
Db 919 GTGACGG 925

Search completed: October 22, 2003, 18:45:04
Job time : 241 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 22, 2003, 11:24:43 ; Search time 47 Seconds
(without alignments)
2044.096 Million cell updates/sec

Title: US-09-747-371-2
Perfect score: 5605
Sequence: 1 MGVAQRNRPQAAWAVLLLLL.....PRSFIRLLRSKVSRLRPYK 999

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 76: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	863.5	15.4	1620	2 T27283	hypothetical prote
2	841.5	15.0	1574	2 T13954	MEGF6 protein - ra
3	727	13.0	2918	2 A54105	fibillin-2 precu
4	724	12.9	2907	2 A57278	fibillin-2 precu
5	691	12.3	2871	2 A55567	fibillin 1 - bow
6	691	12.3	3002	2 A47221	fibillin 1 precu
7	680	12.3	2871	2 A55624	fibillin-1 precu
8	687	12.3	741	2 T46488	hypothetical prote
9	624.5	11.1	1394	2 A35626	transforming grow
10	592	10.6	1712	2 A38261	masking protein pr
11	577.5	10.3	1221	2 A49457	fibulin-2 precu
12	570	10.2	1820	2 A55494	latent transformin
13	567	10.1	685	2 S78040	fibulin, splice fo
14	565	10.1	683	2 C36346	fibulin 1 precu
15	560	10.0	1184	2 A55184	fibulin-2 precu
16	558	10.0	705	2 S34968	fibulin, splice fo
17	553	9.9	601	2 B36346	fibulin 1 precu
18	552.5	9.9	712	2 T42990	fibulin 1, splice
19	540.5	9.6	1251	2 A57293	latent transformin
20	530	9.5	1372	2 T25933	hypothetical prote
21	529.5	9.4	798	2 T22793	hypothetical prote
22	526.5	9.4	2321	2 S78549	notch3 protein - h
23	524.5	9.4	2318	2 A53506	fibulin, splice fo
24	514.5	9.2	689	2 T42760	fibulin-1D precu
25	509.5	9.1	589	2 T43210	notch protein - fr
26	509.5	9.1	2703	1 A24420	hypothetical prote
27	507.5	9.1	3507	2 T34513	notch protein - Af
28	502.5	9.0	2524	2 A35844	notch protein homs
29	497	8.9	2555	2 A40043	

30	494.5	8.8	2531	2 T31070	notch homolog - se
31	493	8.8	2533	2 A49175	Notch B protein -
32	489	8.7	2531	2 S18198	notch protein homo
33	489	8.7	2531	2 A46019	notch-1 protein -
34	487.5	8.7	1964	2 T09059	notch4 - mouse
35	487	8.7	2352	2 T30201	Notch homolog prot
36	484	8.6	2437	2 S42612	transmembrane prot
37	479	8.5	2471	2 A49129	cell-fate determin
38	466	8.3	1064	2 A40136	fibropellin Ia - s
39	443.5	7.9	1429	2 S06434	homeotic protein 1
40	441.5	7.9	1274	2 T42017	cysteine rich prot
41	438	7.8	1217	1 EGVSMG	epidermal growth f
42	431.5	7.7	1133	1 EGR7	epidermal growth f
43	423	7.5	1827	2 T34268	hypothetical prote
44	422	7.5	3623	2 T38618	intrinsic factor-B
45	421	7.5	1207	1 EGRU	epidermal growth f

ALIGNMENTS

RESULT 1
T27283
hypothetical protein Y64G10A.f - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T27283
R:Ainscough, R.
submitted to the EMBL Data Library, September 1999
A:Reference number: 220336
A:Accession: T27283
A:Status: preliminary; translated from GE/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1620 <MIL>
A:Cross-references: EMBL:AL110498; NID:e:542303; PIDN:CA85447.1; CESP:Y64G10A.f
A:Experimental source: clone Y64G10A
C:Genetics:
A:Gene: CESP:Y64G10A.f
A:Introns: 77/1; 116/1; 198/1; 282/1; 365/1; 425/1; 466/1; 549/1; 559/1; 601/1; 625/1

Query Match					15.4%; Score 863.5; DB 2; Length 1620;
Best Local Similarity					24.9%; Pred. No. 1.5e-44;
Matches 280; Conservative 100; Mismatches 372; Indels 371; Gaps 48;					
QY	43	QEDVDECAQGLDCHADALCQNTPTSYKCKPOYQ--GEGROCEDIDECOMELNGGCVH	100		
DB	114	OYDANECAVNNNGGEHE--CUNTIGTYVCRWPGFELSGDGTGSD:DECAVS-NGGCSG	170		
QY	101	DCLNIPGNYRCTCFDGFNLAHDGHNCLDVECLENNGCQHTCVNNKGS--YSCCKKEGF	158		
DB	171	RCVASPSGFRCDPSDLYHADGRTCGKVTSCSTNGSGCEHECENDNGEFYRCRCRVGF	230		
QY	159	FLSDNCHTC-----IHRSEGLSCMN-----KDH	182		
DB	231	KLSENKSCQVPDFCFDNKGGCQHCTNNHRAQCQYGGPHLYDRSRSCVIDECAKYN	290		
QY	183	GCSTHICAPRGSVACRCRPGFELAKNORDC---ILTNHNGGSCQHSCTADGPE-CS	238		
DB	291	GCEHFC-ENKGTVRCRREGYQLGRDGTCEEMLGCGVNGGCGQHDYQDPDGGHVKC	349		
QY	219	CHPQYKMTDGRSLEREDTVLEVTESNTSVVDGDKRVRKRLMETCAV-NNGGCORTC	297		
DB	350	CRNGYILANDQKLC---HDNISTVIHARAPRLWDSYETV-----TCVTFDTLCHKLC	399		
QY	299	KDTSTG-VHSCSPVGFTLQDGKTCKDIDECQTRNGGCDHFCNKIVGSDCGCKKGFLL	356		
DB	400	MHLDSGHVQCCDDGYEL-IDSKFCQDINECHENNGCSQICVNLAGSVECCQCPGFLX	458		
QY	357	TDEKSCQDVDECSLDR-TCDHSCINHPGTAFACNRYGTYLGFTH-CODTNECSINN	414		
DB	459	KDKTCEIDISCSNNNGGCEQICSNQEGGYWCSEPGFELSDGHSCHCNEC:LNNGGC	518		
QY	415	QQVCVNTVGSVECCQHPQYKLNKNCOCV-----EVKGL-----	449		

```

Db      519 AQLCKNRKSGRRCCQFAGYILAHDEKSCVAASDADISNDIEDYKVPGLDSIDEVSS 578
QY      449 ---LPTSVSPR-----VSLHCKKSGGSGCFRLRCHSGIH--SSDVTT 486
Db      579 IESYPADSPRLVFGRRRHVKACVNFQGTLSLELFSSEVRTPDPEKCPNGFFGS--- 634
QY      487 IRTSVTFKLNEKCSLKNALF-----PEG----- 511
Db      635 CQ--SCSDCOQNGKCKMRGSGLLSKDCPSGYGTEKCEQ--CRNGYGVGVCFAKCKCKL--CDP 694
QY      512 ---LRPALPEKSS--VKESF-----RYVNLTCSSGKQVPGAGRPSTPKXFIITVEF 559
Db      695 STGSCREDEPKSDGCPDGPDPGYGSCNLKCRMDCPNGRECDP--VFGYCTCPDGLY--- 748
QY      560 ELSTNQKEVTASCDJSCIVKRTKRLKRAIRLTK-----AVHEEQ 600
Db      749 -----GQCEKPCPHFTFGKNCRFPCKCARENSEGCDEITGKCRCKPGYYGHCKR 799
QY      601 F-----HLQSGNLDVAKKPR-----TSBEQA 624
Db      800 MCSPLGLFAGACWKCSCAPAGIRKCDPVTGDTTKKCPAGYCGNLCDQPCPAGYGVDCQXC 859
QY      625 ESCGVQGHAEHQ-----VSCRAGTY 646
Db      860 SCADVASPHKSKYCHVHTGTCTCLPGKTGLCDQGLIFVETIEFDIAFSINVIACAPNTY 919
QY      647 YDGAERBCILCPNGTQNE--EGMTCEP-----CP--RPG----- 678
Db      920 GPNCATHC--SCVNGAKCDESCHCTPGFYGATCSEVCPTGRFGIDCMQLCKQNGAIC 978
QY      679 ---NSGALKTPKAWNMECCGLCPQGEYSADGAPQOLCALGTF--OPEAGRTSCP----- 730
Db      979 DTSNGSCCAPGSGKXKDKACAPGTFGKDCSKKD--CADGHCDPDSGECICPPKXGH 1037
QY      731 -----CGGSL-----ATKHOGATSFQ---DCETRV-----QCSPGHF----- 759
Db      1038 KCDTCDSLGAGCKGICSCQNGATCDSVTSGCECRPGMKKCDRCPDGRFEGGNA 1097
QY      760 -----YNTTHRC-----IRCPVGYOPEFGKNNC--VSCPQNTTTFDGSIN 800
Db      1098 ICDTTNTDTSYNPVAVCDHVTGECRCPAGWTGPD-----CQTSKP--LGRHGEGRH 1150
QY      801 ITQCKNRRCGGELGDFGTGIESPNYPGNVPANTEC---TWTIN 840
Db      1151 SCQCSN---GASCDRTGCDPCSPGFMKNCSECEPGLWGSN 1190

RESULT 2
MEGF6 protein - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 21-Jul-2000
C:Accession: T13954
R:Nakayama, M.; Nakajima, D.; Nagase, T.; Nomura, N.; Seki, N.; Ohara, O.
Genomics 51, 27-34, 1998
A:Title: Identification of high-molecular-weight proteins with multiple EGF-like motifs
A:Reference number: Z14126; MUID:98360089; PMID:9693030
A:Accession: T13954
A:Status: Preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1574 <NAX>
A:Cross-references: EMBL:AB0111532; NID:G3449293; PIDN:BAA32462.1; PID:G3449294
A:Experimental source: strain Sprague-Dawley; brain
C:Genetics:
A:Gene: MEGF6

Query Match 15.0%; Score 841.5; DB 2; Length 1574;
Best Local Similarity 26.3%; Pred. No. 3e-43;
Matches 245; Conservative 74; Mismatches 289; Indels 323; Gaps 40;

QY      34 .PCGRGAAGPO---EDVDECAQGLDDCHADALQNTFTSKYCKCKEY--QSGECQCEID 88

```

```

Db      109 PCHSKPFGEGCLSDVDEKASANGGC---EGPCNVTGGFYCEC2FGYOI--QSDGKTCQDVD 166
QY      89 ECGNELNGCVHDCNLIPNYRSTCFDGFYLAHDGHNCLDVDELENNGGQOHTCNV--M 147
Db      167 EC--RAHNGGCGHRCVNTPGSYLCECKPGRPRJTDGRTCLAISSTCLGNGGCGHQCVQLT 225
QY      148 GSVECCCKEGFE--SDNQHTCIIHRSERGLSQMKDHGCSHICKEAPRGSAVACECRPGFELA 207
Db      226 TCRCCRCRQYQLQEDGRRCVRRS---PCAEGNGGCKHI COEL--RGLAHCCCHPGYOLA 280
QY      208 XNGRDC--ILTCNNGNGGCGHSCDDTADGPECSCHPQVXVHTDGRSCLEREDTIVLEVES 265
Db      281 ADPKTEDVDECALG--ACCAHACSLNTQSFCKVCVCHAGVELGADGQCRIE---MEIYNS 337
QY      266 NTSVVGDKRYKRRLLNRETCAVNGGCDRTCKTSTGVHSCSPVGTJQLDGKTKCKID 325
Db      338 -----CBAGNGGCSHGSHSTSTPLCTCPRGYELDEQKTCIDID 377
QY      326 ECQTRNGGCDHFCKNIIVSGDCGCKKFKLLJLDEKSCQDVDECSLDR--TCDHSCINHPOT 384
Db      378 DC--ANSPCCQACANTPGGYECSFAGYRLNTDGGCSDVDEKASGHGCGCHSCNLASG 436
QY      385 FACACNRGYT----- 400
Db      437 PCFCCEAGYRLDEDRGCTSLSESVVDLDRUPFYRPLPHIAVLDELPLRPFCDYGAEE 496
QY      401 -----CGD---TNECSI-----NNGG-----CC 415
Db      497 EAAAEALRGHEHTLTKFVCLDSHFHGDLSLTCDDCRNGGTCTFPQDQDCPEGTGII CN 556
QY      416 QVC--VNTVG---SYEQCCHPGYKJHMKKDCVEVKGL--LPTSVSPRVSILR---CGKSG 465
Db      557 ETCPPTTFCKNCSSECTCQ-----NGCTCDPLGACRCPPGVLS---GAHCEDSCP KGF 606
QY      466 GGGCCLRLCHSGIH--SSDVT--RTSVTFKLNKGK--SLKNAELFPEGURLPALPEKHSSVK 524
Db      607 YGKHKRKKCHCA-----NRGRCHRYGACLCDPGLYV----- 638
QY      525 ESFRYVNLTCSSGKQVAGAPGRPTPKEMFIVTEFELETNQKEVTASCDLSIVKRTKR 584
Db      639 ---RPHLACPPWAFPG-----CSEDCLC----- 660
QY      585 LRKAIRTLRAVHREQFHLQLSQMNLDAKPPRTSERQAESCGVCGQHAENQC--VSCRA 643
Db      661 -----EOSH-----TRSCNPKDGSCKKAGFCQGERCAECES 692
QY      644 GTYYGARERCIL-----CPKSTF--QNEGGQVTC-- 671
Db      693 GFFPGCRHRCCTCFGVACDPVSGECRTQCQPPGYGEGDGGCECPVGTFGVNGSGSCSVG 752
QY      672 EPCPR-----PNSGALKJTPKAWNXXSECGLCPQGEYSADGAPQOLCALG--TFQPE 722
Db      753 APCHRVYTGELCFPPGVTG-----EDCGADCPBGRKSGCGCEI CPACENHASCNPE 802
QY      723 AGTSFCPPCGGLATKHQCATSF---QDCETRVCCSPGHFYNTTHRCI----- 768
Db      803 TGTCLCLP--GFVGSRCQPTCSAGYGTGTCQIRACANSDGHCDPTTGRCSAPGWTG--SC 860
QY      769 --ACPUGTYQPE-----FGKNNVCVSCPG 789
Db      861 QRACDSGHGWGPDCLHPCNCSAGHGRKCDVSG 891

RESULT 3
A541C5
fibrillin-2 precursor - human
C:Species: Homo sapiens (Man)
C:Date: 06-Sep-1994 #sequence_revision 09-Sep-1994 #text_change 02-Aug-2002
C:Accession: A54105; S17063; S31101
R:Zhang, H.; Apfelroth, S.D.; Hu, W.; Davis, E.C.; Sanguinetti, G.; Bonadio, J.; Necham
C:Cell Bio. 124, 955-963, 1994
A:Title: Structure and expression of fibrillin-2, a novel microfibrillar component pte
A:Reference number: A541C5; MUID:94165150; PMID:9:201C5

```

[illegible]

```

1650 IDECELPGLCGGNCCI-----N 1667
QY 62: ERQAESGVGGGH--AE-----NOCVS---CRAGTYDGARERCILCP----- 658
Db 1668 TFGSFQCECPQGYLVSEDTKED:DECFAPHPGVCGPOTCYNTLGNVTCICPPEYMQVNG 1727
QY 659 -----NG-TFQNE-----EQMTTCBCPDRPGNSG-- 661
Db 1728 GHNCVDMKSFYRSYNGTTCENSELPFNVTKRMCCCTYVKGAKGKPCPCPTPGTADFK 1787
QY 682 -----ALKTTEAMNSEC----- 694
Db 1788 TICGNIPGFTDIHTGKAVDIDECKEIPGICANGVCINQIGSFCECPTGFSYNDLLVC 1847
QY 695 -----GSQQ-----PGEYSADGPAPQLCALG-TFOEAG---RTSCP-- 730
Db 1848 EDIDECSSNGDNLCQRNADCIINSGSYRCE-----CAAGFKLSPGACVDRNECEIPN 1900
QY 731 -CGGGLATKHGATSFQ-DCSTRVQCSFGHYNTTHRCIR--CPVGTGVEFGKNCVSV 786
Db 1901 VCSHGLCVDLQG--SYGICIHNGFKASQDTMCMDVDECEHPFCGNGTCTNTVGSYNCLC 1958
QY 787 CGP-----NTTIIDFGSTNI--TQCKNRRCGGELGDF-----TGYIESPNYPCNYPANT 833
Db 1959 YPGFELTHNDCLEDECCSFQGVGRNRCFNEIGSFKCLNESYELTIDGKNKICITN- 2017
QY 834 EC 835
Db 2018 EC 2019

RESULT 4
A57278
fibrillin-2 precursor - mouse
C:Species: Mus musculus (house mouse)
C>Date: 23-Feb-1996 #sequence_revision 23-Feb-1996 #text_change 02-Aug-2002
C:Accession: A57278
R:Zhang, H.; Hu, W.; Ramirez, F.
J. Cell Biol. 129, 1165-1176, 1995
A:Title: Developmental expression of fibrillin genes suggests heterogeneity of extrace
A:Reference number: A57278; MJD:95263670; PMID:7744963
A:Accession: A57278
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2997 <ZHA>
A:Cross-references: GB:L39790; NID:G762810; PID:AAA74908.; FID:G762811
C:Superfamily: fibrillin 1; EGF homology
F:1239-1274/Domain: EGF homology <EGF>
F:2488-2523/Domain: EGF homology <EGF>

Query Match: 12.9%; Score 724; DB 2; Length 2907;
Best Local Similarity 26.1%; Pred No. 7.7e-36;
Matches 25.; Conservative 98; Mismatches 300; Indels 3.; Gaps 51;

QY 45 DVECAQGLDCHADALCQNTPTSYKSCPKGYQG--EGROCEDIDCGNELNGGCVHDC 102
Db 1193 DINECSLS-DMLCRNGKCVNMIGTVQSCNPGYQATPRQCTDIDEC-WIMNGGCDTQC 1250
QY 103 LN-PGNYRCTCFDGFMLAHGHNCLDVIDCLENN-----GCQHTCVNVVMSYECCCKEG 157
Db 1251 TNSGSEYSCSGEYALMPDRSCADIDEC-ENNPDI CDGG--QC-NIPGEYRCLCYDG 1306
QY 158 PFLSNQHTCIHRSBEGLSMKNKHGSHICK----EAPRGSVACERCPGPELAKNRDC 213
Db 1307 PNASMDVXTCIDVNECDLN-----PNICYGECENTKGSFICHQQLGYSVVKKGTGC 1358
QY 214 --ILTCNHNNGGC--QHSCTDADPGCSCHPQYVMMHTDGRSCLEREDTVLVETSKTS 269
Db 1359 TDVDECEI GAHNCMDHASCNLVPSFKCSREGWV--GNGIKCIDJDEC-----ANGT- 1409
QY 270 VDDGKRVKRLJLMEVCANVNGGCDRTCKTSTTVHSCSPVGFTQLDGGTKCKIDECQT 329
Db 1410 -----HQCSN-----AOCVNTPGSYRCACSEGT--CGGTCGSDVDECAE 1448

```

```

QY 330 RGGCGDH-FCXNIVSGFDGCKGKFKLLTDBKSCQDVDECSLRTCDH-SCINHPGTAC 387
DB 1449 NTNLCENGOC-LNPGAYRCECEMGFTPASDRSCQDIDECSPQNICVFGTCNN-LPGMFHC 1508
QY 388 ACNRGYTL-YGTHCGDTNECS--IN--NGGQQCVVNTVGSYEQCHPGYKHLHNNKDC 442
DB 1509 ICDDGYELDRGTGNTDIDECADPINCNG--LCVNTPRYECNCPFPCLNATGVC 1564
QY 443 VEVK-GLLPSTVSPVSLHCGKSGGDCGFLCHSGIHLSSDVTIRTSVTFKLNEKCS 501
DB 1565 VZNRVNCYLKTPGR-----GDGS-LSCNT---EAGVGVSRSOC-----CS 1662
QY 502 LKXAEFLP-EGRLPALPEKHSSVKESPRYVNLTCSSGQVPGAGRPSTPKEMFITVEFE 560
DB 1603 LKXAGNCPETCPPNNSHTYTL-----CPGEGFRNP-----ITILE 1642
QY 561 LETNOKEVTASCD-LSCIVKRTKRLRAIRTLKRAVHREQPHQLSGMNLVAKKPPRT 619
DB 1643 DIDECELPGLCOGNCI-----1660
QY 620 SERQBSQVGGGH--AE-----NQCVS-----CRAGTYDGAARECILCP-----658
DB 1661 NTFGSQCECPQGYLSEBTRICEIDECFAHPGVCGPCTVNTLGNVTCICPPEYMOVN 1720
QY 659 -----NG-TFONE-----EGQMTCEPCPRGRSG- 681
DB 1721 GGHNCMDMRKSCFYRSYNGTTTCEBELPENVTRMCCCTYNVKGAKNCPCEPCTGTADF 1780
QY 682 -----ALKTPAMNSEC-----694
DB 1781 KTCIGNIPGFTDIHTGRAVDIDECKEIPGICANGVCINQIGSPCECPTGFSYNDLLV 1840
QY 695 -----GGLCO-----PEYSGDGFAPQLCALG-TFQFAG---RTSCPP-- 730
DB 1841 CEDIDECNSGNDLCORNADCNPSGYRCE-----CAAGFKLSPGACVDRNECLEIP 1893
QY 731 ---CGGLATKHOGATSFQ-DCETRVCQSPGFHNTTHRCIR--CPVGTQYQPEFGONCV 785
DB 1894 NVCSHGLCVDLG--SYQICLNGFKASQDQYCMQDVDECEHPCANGCTCKNTVGSYNCL 1951
QY 786 SCGP-----NTTDFDGSNTI--TQCKNRRCGGELGDF-----TGYTESPNYPGNYPAN 832
DB 1952 CYPGPELTHNNDCLDIDECSPFGVCGRNCRFNEIGSKLCLNEGELTDPGKNCIDTN 2011
QY 833 TEC 835
DB 2012 -EC 2013

RESULT 5
A:55567
C:Species: Bos primigenius taurus (cattle)
C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 02-Aug-2002
C:Accession: A55567
R:Filetra, D.J.; Li, L.; Potter, K.A.; Womack, J.; Byers, P.H.
Genomics 23, 480-485, 1994
A:Title: Sequence of the coding region of the bovine fibrillin cDNA and localization to
A:Reference number: A55567; MUID:95137597; PMID:7835900
A:Accession: A55567
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-2871 <TIL>
A:Cross-references: GB:L28748; NID:9508427; PIDN:AAA74122.1; PID:9508428
C:Superfamily: fibrillin 1; EGF homology
F:1201-1236/Domain: EGF homology <EGF>

Query Match 12.3%; Score 691; DB 2; Length 2871;
Best Local Similarity 22.7%; Pred. No. 7,6e-34;
Matches 271; Conservative 107; Mismatches 375; Indels 440; Gaps 49;
QY 45 PVDECAQGLDDCHADALCONTPTSYKCSCKPGYQGEGRCCEDIDECNELNGGCVHDCLN 104

```

```

DB 1363 DLEBSNGTHMSCADCKNTMGSYAC-LKEGYTGTGTTCTDCLDECSNMLNCGNGOCLN 1422
QY 105 IPONVACTCFDGFYLAHDGHNCLDVEEC-----LENNG 137
DB 1423 AFGGYRCECDMEFVPSADGACEDIDECSLPNICVFGTCHNLPGLPFCECISGVELDRSG 1482
QY 138 G-----CQHTCUNMGSYECCKEGEFLSONQHTCIH-----RS 171
DB 1483 GNCIDVNECLDPTTCISGNCVNTPSYTCDCPPDFELNPTRVGCVOTRSNGCYLDIRPRG 1542
QY 172 BEG-LSCMN-----KDHGCSHI-----187
DB 1543 DNGDTACSEI-GVYSKASCCSLKAWTPELCPVPNTSEYKILCPGGEGFRNPITV 1602
QY 188 -----CKEAPR-----GSVACEAPGFELAKNQDCC-IJTCNH-----GNNG 223
DB 1603 ILEDIDECQELPGLCOGKGCINTFGSFQRCPTGYLYNEDTRVCDVNECETPGICGEG- 1661
QY 224 CQHSDDTADGPECSCHPOYKMHITGRSCLEREDTVLEVTESNTTSVVGOKRVRKRLM 283
DB 1662 ---TCVNTVGNVTCICPPDYMQVNGNNDNRSLCYRNYAQNQTCDGELLFNMTKMM 1718
QY 284 ETCVANNNG-GCDRTCK-----DTSTGV-----304
DB 1719 CCOSYNIGRAMNKPCEQCPIPTSTDEPATLGSQRPGFVIDIVTG-LPVDICEKEIFGVCE 1778
QY 305 -----HCSQVGVFTLQDKTKCDIDECG-----328
DB 1779 NGVCINMGVSFRCEPVGFFYNDKLLVCEIDECQNGPVCORNAECINTAGSYRCDCKPG 1838
QY 329 ---TRNGGCDH-----FCXNIVSGFDGCKGKFKLLTDBKSCQDVDECSLD 371
DB 1839 YAFSTGQCNDRNECEGFENICSHGCCIDTVGSFVCLHTGFKTNADOTMCLDINECEPD 1898
QY 372 RTCDHSCINKHPTFACACNRGYTYGFTHCGTNECSINNGG-CQQ-VCVNTVGSYEQC 429
DB 1899 ACNGTGRNTIGSFNCRNNGHFIILSHNNDICIDVDECATGNLNCNGQCCINTVGSFQCC 1958
QY 430 HGYKYLHNNKCKCEVVKGLJ---PTSVSPVRS-JHCKSKGGGDCGF-PCHSIHLSSDVTI 487
DB 1959 NEGVEYAPGRTCVQINECLDPRKCAPGT---CNLDGSYRCI--CPFGVSLQNDKCE- 2012
QY 488 RSVTEKLNE-----GKCSLKNAE-----JFPEGLR-PALPEKHSSVKESPRYVN 531
DB 2013 -----DIDECVESPEICALGTCS--NTGSKFCLCPGFSLSSTGRCDLRMSYCAK 2064
QY 532 LTCSSKQVPGAPRPSTPKEMFITVPELETNQKEVTASCDLS-----CIVKATEKR 584
DB 2065 F-----EGGKCSPK-----SRNHSKQECCLXGEGMGDPCELCPTED- 2103
QY 585 LAKAIRTLKRAVHREQFH-LQL-----SGXKLDVAKKPERTSERQAES-----626
DB 2104 -----CPAPRQICPYSGIIVGDDDAVDMDECKEPDVCCKHGGC-WTDGSY- 2149
QY 627 ---CGVGQGHAEQCV---SCRAGTYDGAARECICPNNGTFSQNEEGMTCPCPRPGNS 680
DB 2150 RCSCPFGYLQGNCEVDTDECSVGN-----PCGKGTCKNVIGGFEC-TCEEGEP 2198
QY 681 GALKTPENANMSECGG---LCQPGEYSDGFPAPQLCALGTQFPAGRTSC---PFCGGG 734
DB 2199 GPWMTCE--DINECAONPLLCAPRCVNTVGSYECK-CFAGYVYLR-DRNMCKDEDECEG 2254
QY 735 LAKHOGATSFDCEPRVQ-----CSPGFHYNTTHRCIRCPVGTYQPEFGKNCNVSPG 789
DB 2255 ---KHDCAEKQNECKULICTYLCICGPGYQRPDPDEGCV-----DNECQTAPG 2300
QY 790 NITTFDGSNTITQCKNRRCGGELGDF-----GYIESPNYPGNYPANTECTWTINPFPK 844
DB 2301 -----ICENGRCLNTRSYTCECNDGFTASN-----2327
QY 845 RRLILVPELFIPLPDDCCG-----YLVNRKTSSSNSVTTTETCQTYER-----888

```

Db 2328 -----QDECLNREGYCFTEVLQMCQIGSSNRPVTKSECCCGRGWGP 2373

Qy 889 -----PIAFTSRSKLWQFKSNEGSARGFQVPVTVYDEYQELIEDIVRDG 936

Db 2374 HCEICPFQQTAFKLC-----PHGRGFMTNGADIDE--CKVIHVDVCRNG 2416

RESULT 6

A47221

fibrillin 1 precursor - human (fragment)

C:Species: Homo sapiens (man)

C:Date: 02-Jun-1995 #sequence revision 25-Apr-1997 #text change 02-Aug-2002

C:Accession: A47221; 154355; S17064; 159574; S17062; S6211; A34198

R:Corson, G.M.; Chabberg, S.C.; Dietz, H.C.; Charbonneau, N.L.; Sakai, L.Y.

Genomics 17, 476-484, 1993

A:Title: Fibrillin binds calcium and is coded by cDNAs that reveal a multidomain structure

A:Reference number: A47221; MUID:94010947; PMID:7691719

A:Accession: A47221

A:Molecule type: mRNA

A:Residues: 1-337, 'T', 339-1029 <COR>

A:Cross-references: GB:X63556

R:Pereira, L.V.; D'Alessio, M.; Ramirez, F.; Lynch, J.; Sykes, B.; Pangilinan, T.; Bonad

Hum. Mol. Genet. 2, 961-968, 1993

A:Title: Genomic organization of the sequence coding for fibrillin, the defective gene p

A:Reference number: 154355; MUID:93372860; PMID:8364578

A:Accession: 154355

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 132-3202 <PER>

A:Cross-references: GB:L13923; NID:9306745; PIDN:AAB2036.1; PID:9306746

R:Maslen, C.L.; Corson, G.M.; Maddox, B.K.; Glanville, R.W.; Sakai, L.Y.

Nature 352, 334-337, 1991

A:Title: Partial sequence of a candidate gene for the Marfan syndrome.

A:Reference number: S17064; MUID:91304568; PMID:1852207

A:Accession: S17064

A:Molecule type: mRNA

A:Residues: 1030-3002 <MAS>

A:Cross-references: EMBL:X63556

R:Dietz, H.C.; Valle, D.; Francomano, C.A.; Kendzior, R.J.

Science 259, 680-683, 1993

A:Title: The skipping of constitutive exons in vivo induced by nonsense mutations.

A:Reference number: 159574; MUID:93157831; PMID:8430317

A:Accession: 159574

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 2217-2288, 'I', 2290-2325 <RES>

A:Cross-references: GB:S54426; NID:9264860; PIDN:AAB25244.1; PID:9264861

R:Lee, B.; Godfrey, M.; Vitale, E.; Horst, H.; Mattei, M.G.; Safarazi, M.; Tsipouras, P.

Nature 352, 330-334, 1991

A:Title: Linkage of Marfan syndrome and a phenotypically related disorder to two differ

A:Reference number: S17062; MUID:91304567; PMID:1852206

A:Accession: S17062

A:Molecule type: mRNA

A:Residues: VLVTVWFIFSYNKM'L, 944-1444 <LEE1>

A:Cross-references: EMBL:X62008; NID:931398; PIDN:CAB56534.1; PID:95924015

A:Accession: S62111

A:Molecule type: protein

A:Residues: 1166-1176, 'X', 1178-1180, 'D', 1182-1185 <LEE2>

R:Maddox, B.K.; Sakai, L.Y.; Keene, D.R.; Glanville, R.W.

J. Biol. Chem. 264, 21381-21385, 1989

A:Title: Connective tissue microfibrils. Isolation and characterization of three large p

A:Reference number: A34198; MUID:90078246; PMID:2512293

A:Accession: A34198

A:Molecule type: protein

A:Residues: 565-575, 1890-1892, 'I', 1894-1900 <MAD>

C:Comment: Fibrillin is a major component of elastin-associated microfibrils.

C:Genetics:

A:Gene: GDB:FBN1

A:Cross-references: GDB:127115; OMIM:134797; OMIM:154700

A:Map position: 15q21.1-15q21.1

A:Mutons: 2236/1; 2258/1; 2297/1

C:Superfamily: fibrillin 1; EGF homology

C:Keywords: alternative splicing; calcium binding; extracellular matrix; glycoprotein; M

F:13062/Product: fibrillin (5'-region exon A splice form) (fragment) #status predicted
 F:132-3002/Product: fibrillin (5'-region exon C splice form) #status predicted <MAD>
 F:1332-1367/Domains: EGF homology <EGF>
 F:1457-1492/Domains: EGF homology <EGF2>
 F:2262-2295/Domains: EGF homology <EGF1>

Query Match 12.3% Score 691; DB 2; Length 3002;

Best Local Similarity 22.5% Pred. No. 8e-34;
 Matches 269; Conservative 105; Mismatches 378; Indels 442; Gaps 45;

Qy 45 DVDECAOGLDCHADALCCQNTPTSYKSCCKPGYQGEORCEDIDEGNEINLGGCVHDCLN 164

Db 1494 DLDECSNGTMCQSHADCKVTMSYRLCKEGYTGDFCTDLDSENLKLCNGOCLN 1553

Qy 105 IPGNRYCTCFDGYLAHDHNCLEVBECLENNGGQCHTCVNVMSYECCKEGLFSDNQ 164

Db 1554 APGYRCECDMGFVPSADGKACEDIDECSLPNICVFCTCHLPSLFRCECEIGVELDRSG 1613

Qy 165 HTCHRSSEGLSCNKHGSHICKAPROSVACECEPGFELAKNQRDCILT----- 216

Db 1614 GNCTDYNE-----CLDPTTC:SGNCVATP-GSY:CCGPPDFELNPTRVGCVTRSGCYLD 1668

Qy 217 ----- 216

Db 1669 IRPRGNGDFACNEIOGVSKASCCCSLKANGTPECMCFAYNTSEYKLCRSGEGFRP 1728

Qy 217 -----CNHSGNGGCG----- 233

Db 1729 NPITVILEIDICEQELSLGCGGGKICNTFGSFQCRCPYGVYLNEDTRVCDVNECEPGL 1788

Qy 234 -GP-----BSCHPQYKHGTGRSCLEREDTVLEVTESVTVSGDKVAKRLJ 282

Db 1789 CGPCTCVTYGNTYCTICPPDYMQVNGGNCKMRRSLCYRYADNCTCGELLFNMTXK 1848

Qy 263 YETCAVANG-GCGRCK-----DTSTGV----- 304

Db 1849 MCCSYNIGRANKPCECCIPSTDEPATLCGSGRPGFVIDYVGLPVIDEGREIPGVC 1908

Qy 305 -----HGCSPVGTQLDGRTKDIDECQ----- 328

Db 1909 ENGVCINWVSFRCECPVGFYNDKLLVCEIDECQNGPVCQNAECINTAGSYRCCKP 1968

Qy 329 -----TRNGGCGH-----PCNIVSGFDGCKKGFALLTDEKSCQDVDECSL 370

Db 1969 GYRFTSTGQCKDRNECEIFNICSHGQCIDTVGSFYCLHTGKTNDDCTMCLDINCEK 2028

Qy 371 DETCDHSCINHPGTFACACNRYGTYLGYTHCGGTNECSINNGG-CQG-VCVNTVGSYEQ 428

Db 2029 DACNGTCRNTIGSFNCRCHGFFLSHNNDCLDVDECSNGNLCRNGQCINTVGSFQCG 2088

Qy 429 CHPGYKLFHWKDKCVYKGLL--PTSVSRVSLHCGKSGGGGCGFLRCHSGIHLSDDVT 486

Db 2089 CNEGYEAPDORTCVINECLLEPRKCAPGT---CONLDGSYRCI--CPGYSLQNEKE 2143

Qy 487 IRTSVTKLANE-----GKCSLQAE-----LPPEGLR-PALPEKISSVKEFRIV 530

Db 2144 -----DIDECVEPEICALGTCS--NTSGSFKCLCPGEGFSLSSSGRRQCDLRMSCYA 2194

Qy 531 NLTCSGKQVPGAPGRESTPKEMPIITVEFELETNKEVTASCDLS-----CLVKEKTEK 583

Db 2195 KF-----EGKCSSPK-----SRNHSKQECCLAKGSGWGFDELCCTEP 2234

Qy 584 RLKRAIRLTKAVTHREQFHLQL-----SQNLDVAKKPPRTSERQAES----- 626

Db 2235 -----DEAFQCPYSGSIIIVGPDSDAVDVECKECPVCKHGQCIINTDGS 2279

Qy 627 ----CGVQGHAEHNOCV---SCRAGTYDGAERECILCPNGTTPONEGQVTCPCPRPGN 679

Db 2280 YRCECPFGYTLAGNECVTDECSVGN-----PCGNGTCKNV-GGFEC-TCEEGFE 2328

Qy 680 SGALKTFEAMNMSECG-----JCQPGEVSADGAPCQCALCJTFQEGAGRTSC---FFCGG 733

Db 2329 PGPMWTCE--CINECAQNPLLCAPFCVNTYGSYECK-CPVGYVLFE--DRMCKDEDECEE 2394

```

QY 734 GLATKHQGAISFDQCETRVQ-----CSPGHFVNTTHRCIRCFVGTQYQPFSGKNKNCVSCP 788
DB 2385 G---KHDCTEKQMBCKNLTGYTMCICGFGYQRRPDGEGCV-----DENECQTKP 2430
QY 789 GNTTDFDGSNTITCKNRRCGELGDF-----GYIESPNYPGNYPANTBCTWTIRPPP 843
DB 2431 G-----ICENGRCLNTGSYTCRCNDGFTASPN-----2456
QY 844 KRRILIWPFEIFLPEDDCGD-----YLWRAKTSSNSVTTTTCQTYER--- 888
DB 2459 -----QDECLDNREGYCFTEVLQNMCIQSSNRNFTVSKSECCCGGRGWG 2503
QY 889 -----PIAFTSRKSLWTFQKSNBNSARGQVFPVYDYDEYQELIEDIVRDG 936
DB 2504 PHCEICPFQGTVAFKKLC-----PHGRGFTMGADIDE--CKVIHDVCRNG 2547

RESULT 7
A55624
Fibrillin-1 precursor - mouse
C:Species: Mus musculus (house mouse)
C:Date: 23-Mar-1995 #sequence_revision 23-Mar-1995 #text_change 02-Aug-2002
C:Accession: A55624
R:Vin, W.; Smiley, E.; Germlinger, J.; Sanguinetti, C.; Lawton, T.; Pereira, L.; Ramirez,
J.; Biol. Chem. 270, 1798-1806, 1995
A:Title: Primary structure and developmental expression of Fbn-1, the mouse fibrillin gene
A:Reference number: A55624; MUID:95130561; PMID:7829516
A:Accession: A55624
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-2871 <YIN>
A:Cross-references: GB:L29454; NID:9575509; PIDN:AAA56840.; P:ID:957555:0
C:Genetics:
A:Gene: Fbn-1
A:Superfamily: fibrillin 1; EGF homology
F:1201-1236/Domain: EGF homology <EGF>

Query Match 12.3%; Score 690; DB 2; Length 2871;
Best Local Similarity 23.4%; Pred. No. 8,8e-34;
Matches 276; Conservative 104; Mismatches 384; Indels 418; Gaps 51;

QY 45 DVDECAQGLDDCHADALCONTPTSFKSCCKPGYQGEGRQCEIDECGNELMGCVHDCLN 104
DB 1363 DLDCEGNGTHMCSQADCKNTWGSYRCLCKDGTGGFTCTDLDCSENLKLCGNQCCLN 1422
QY 1205 IFGNVRCCTCFDGFMLAHGHNCLDVDEC-----JENNG 137
DB 1423 APGGYRCCECMGFVPSADGKACEDIDECSLPNICVFGTCHNLPLFRCECEIGYELDRSG 1482
QY 138 G-----CQHTCVNMGSYECCKEGFFLSDNQHTCIH-----RS 171
DB 1483 GNCITVNECLDPTTCISGNCVNTPGSYTCDSPDFELAPTRVGCVDTRSGNLYNIRPG 1542
QY 172 BEGLSOMXNDHG-----CSHI-----187
DB 1543 DNGDTACSNIEIGVGSKASCCSLGKAWCTPCELCFSVNTSEYKILCPGGEGRFNPIIV 1602
QY 188 -----CKEAPR-----GSVACEGRFGEJAKNQDCC--ILTNH-----GNNG 223
DB 1603 ILIEDICQELPGLCGGKCIINTFGSFQCRCTPGYLYNEDTRVCDVNECECTPGICGPG- 1661
QY 224 CQHSCDDTAGDPCSCCHPOYKMTDGRSCLEREDTVLEVTESNTTSSVVDGDKRVKRRJLM 283
DB 1662 ----TCVNTGVNTYTCIFPDYMQVNGNCKMDRRSICRYNADNCTCDGELLFNNTKM 1718
QY 284 ETCAVN-----NGCCORTCKDTST-----302
DB 1719 CCSSYNIRRAWNKPCEQ-CP1PSTDEFAFLCGSQRPGFVIDIYTLGLPVDIDECREIPGVC 1777
QY 303 --GV-----HCSPGVFTLQDKTKDIDECO-----328
DB 1778 ENGVCINMGVSRCECFVGFYYNDKLLVGEDIDECQNGPVCLRNAECINTAGSYRCDCKP 1837

```

```

QY 329 -----TRNGCCH-----FCNIVGSEFDCGCKKGFKLLTDEKSCQVDDESL 370
DB 1838 GYRLTSTGQCNDRNECOEIPNICSHGQCIDTVSSFYCLHTGFKTNEQTVCLDNESE 1897
QY 371 DRFCDSHSCINHGTTPACACNRGYTLGYFHCQGTNECSINNG-CCQ-VCVNVTGSECO 428
DB 1998 DACNGTCTCNTTGSFYRCRCNHGFIILSHNDCIDVCECATGNLJCRNGOCVNTVGSFQR 1957
QY 429 CHPGYKLRNKKDCVEVKG--LTPSVSPRSLHCKSGGGGDCFLRHSGLHLSDDVT 486
DB 1958 CNEGYEVAEDGRCTVDINECVLDPGKCAPGT---CONLDGSYRC--CPPGYSLQNDKCE 2012
QY 487 IRTSVTFKNE-----OKCSKNAE-----LFPEGLR-PALPEKHSVSKSPRYV 530
DB 2013 -----DIDECVEEPEICAJGTCS--NTEGSFKCLCEGFSWSSSGRRCCQDRMSYCYA 2063
QY 531 NLTCSGKQVPGAPGRPSTPKEMFITVFELETNKEVITASCDLS-----CIVKRTKX 583
DB 2064 KP-----EGKCSSPK-----SRNHSKQECCLALKEGWDGDCPELCPTEP 2103
QY 584 RLKKAIRTLRKAVHREQFHLQ-----SGMNLQVAKKPP-----RTSER 622
DB 2104 -----DEAPRQICPFGSGIIVGPDSDAVDMDECKEPPDVCRHGQCINTGGS 2148
QY 623 QAESCGVGQGHAEHQCV---SCRAGTYDGAERECILCPNGTFFQNEEGOMTCEPCRPON 679
DB 2149 YRCECPFGVILEGNECVDTDECSVGN-----PCGNGTCKXVIGGFEC--EEGEFE 2197
QY 680 SGALKTPPEANMSECGG---LQCPGEYSADGFAPQCALGTQFQPEAGRTSC---FPCSG 733
DB 2198 PGPMTCE--DINECAQNPJLCAFRVNTYGYECK-CPGVYVJRE-DERMCKDEDECAE 2253
QY 734 GLATKHQGAISFDQCETRVQ-----CSPGHFVNTTHRCIRCFVGTQYQPFSGKNKNCVSCP 788
DB 2254 G---KHDCTEKQMBCKNLTGYTMCICGFGYQRRPDGEGCV-----DENECQTKP 2299
QY 789 GNTTDFDGSNTITCKNRRCGELGDF-----GYIESPNYPGNYPANTBCTWTINIRPP 843
DB 2300 G-----ICENGRCLNTGSYTCRCNDGFTAS-----PTODECL-----2332
QY 844 KRRILIWPFEIFLPEDDCGDYLWRAKTSSNSVTTTTCQTYER-----PIAFTS 894
DB 2333 DNEGYCFSEVP--ENMC-----QGSNRPVTKSECCVGGRGNGLHCEICPEGTV 2384
QY 895 RSXKLWIOQKNEGNSARGQVFPVYDYDEYQELIEDIVRDG 936
DB 2385 AYKKLC-----PHGRGFTMGADVDE--CKVIHDVCRNG 2416

RESULT 8
T46488
Hypothetical protein DKFZp434J065.1 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
C:Accession: T46488
R:Ducrest-Hoeft, A.; Lauber, J.; Yewes, H.W.; Gassenhuber, J.; Wienann, S.
submitted to the Protein Sequence Database, January 2000
A:Reference number: Z23035
A:Accession: T46488
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-741 <AAA>
A:Cross-references: EMBL:AL237638
A:Experimental source: adult testis; clone DKFZp434J065
C:Genetics:
A:Note: DKFZp434J065.1

Query Match 12.3%; Score 687; DB 2; Length 741;
Best Local Similarity 31.7%; Pred. No. 3.3e-34;
Matches 152; Conservative 61; Mismatches 196; Indels 70; Gaps 15;

QY 37 GRAAGQEDVDE--CAQGLDCHADALCONTPTSFKSCCKPGY--QGEGRQCEIDECGN 92

```



```
Db      20 GEAKRTRKTLBHCAH-----FCINIPGSVYCRCKGKIYILNSDQTTCTRIODLCAM 70
QY      93 ELNGGCVHDLNIPGNRYRCTCPDGFMLAHGHNCLDVEDLENNGGCQHTCVNMGSYEC 152
Db      71 E-DHNCQELCVNVPGSFVQCYSYALAEDEGKRCVAVDYCASENHGCBHECVNADGSVLC 129
QY      153 CKKEGFFLSDNHTCIHRSEBGLSCMKNDHGSJCKEAPRGSVACRCRPFSEIAKQORD 212
Db      130 QCHEGFALNPDEKTKTKID----YASSNHGQCEHCVNTD-DSYSCHCLGKFTLNPDKKT 184
QY      213 C--ILTCNHGNGCGHSCDDTADGPECSCCHPQKMHKTDCRSCLEREDTVLEVTESNTTSV 272
Db      185 CRINVCALNKPGECHECVNMBSYVCRHRYTIDPQNKCSR----- 228
QY      271 VDGKRVKRRLLMETCAVNGGCDRTCKDTSTGVHSCFPVGFRTLCQDGKTKDIDECQTR 330
Db      229 -----VDHCAQDHGCEQLCLNTEDSFVQCSEGF--NEDLTKTSRVVYCLLS 276
QY      331 NGGCDHFCNIVGSDPGCKGFKLLTDEKSCODVDEGL-DRTCCHSCINHPGTFACAC 389
Db      277 DHGCEYSVCNMDRSFACQPEGHVLRSDGKTCAKLDSALGSHGCEHSCVSSSEDSFVQCQ 336
QY      390 NSGYTLY--GFTHCGDTRFCS--NGGCGQVQVNTVGSVECCCHPCYKLNKMKDCEVVKG 447
Db      337 FEGYILREDGKT-CRRKQVQIDHGCEHICVNSDSTYCELGFRLAEDGKRC-RRRD 394
QY      448 LLPTSVSPVSLHCGKS-----GGGDCGFLRCHSG:HLSSDVTTRTSTVTFKLNESKCSL 502
Db      395 VCK-----STHGGCEHICVNGNSYICKGSEGFVLAEDGRCK-----KCTEGP:DL 441

RESULT 9
A35626
transforming growth factor beta-1-binding protein - human
C:Species: Homo sapiens (man)
C>Date: 21-Sep-1990 #sequence_revision 21-Sep-1990 #text_change 11-Jan-2000
C:Accession: A35626
R:Kanzaki, T.; Olofsson, A.; Moren, A.; Wernstedt, C.; Hellman, U.; Miyazono, K.; Claess
Cell 61, 1051-1061, 1990
A:Title: TGF-beta1 binding protein: a component of the large latent complex of TGF-beta1
A:Reference number: A35626, MUID:90275601, PMID:2350783
A:Accession: A35626
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-1394 <KAN>
A:Cross-references: GB:M34057; NID:G339547; P:DN:AAA61160.1; PID:G339548
C:Superfamily: unassigned EGF-related proteins; EGF homology
C:Keywords: alternative splicing
F:750-791/Domain: EGF homology <EGF>

Query Match 11.1%; Score 624.5; DB 2; Length 1394;
Best Local Similarity 24.9%; Pred. No. 3.9e-30;
Matches 240; Conservative 94; Mismatches 334; Indels 297; Gaps 48;

QY      45 DVDECAQGGDDCHADALCONTTSYKSCPKGYQ--GEGRCQEDIDECNENLNGCVHD- 102
Db      546 EINECTVNPDIQGA-GHCINLPVRYTCICYEGRFSEQGRKCVDDDEC-TQVQHLC:CSQGR 603
QY      102 CLNIPGNRYRCTCPDGFMLAHGHNCLDVEDLENNGGCQHTCVNMGSYEC-CCKEGFFL 160
Db      604 CENTGESFLICIPAGPMASEEGTNCIDVDECLRPDVCGEHCVNTVAFREYCDSGYRM 663
QY      161 SNQHTCIHRSEBGLSCMKNDHGSJCKEAPRGSVACRCRPFSEIAKQORDCILTCNHG 220
Db      664 TOR-----GRCEIDEDCLNPSTCPDQCVNSP-GSYQC-----VPC:EG 701
QY      221 NGGCGHSCDDTADGPECSCCHPQKMHKTDCRSCLEREDTVLEVTESNTTSVVDGDKRVKRR 280
Db      702 FRGWNQOCLDVDE-----CLE----- 717
QY      281 LLMETCAVNGGCDRTCKDTSTGVHSCVPGVFTLO:DGKTKCIDECQTRNGGCDHFCN 340
```

```
Db      718 --PNVCA--NG-----DCSNLEBSYMCSCHKGVTRTPDHKGHCRDIDECQCGNLCVNGCCKN 769
QY      341 IVGSFDCGCKKFKLLTDEKSCQDVDECSLDRTCDH-SCINHPGT:FAACNRGYTYXGF- 398
Db      770 TESSFRCTCGQYQLSAAKQDCED:DECQHRHLCAHQCRNTEGSCFQCVCDGYRASGLG 829
QY      399 THCGDTRNECSINNGGCGQ--VCVNTVGSYECQCHPGYKLNHWK-----KDCVEVKGLPPTS 452
Db      830 DHCEIDNECLDKSVQCRGDCINTAGSYDCTCPDQFQ:LDNKTQCDINBC-EHPGL----- 864
QY      453 VSPRVLHGKSG--GGDGF-LRCHSGIHLSSDVTTRTSTVTFKLNESKCSLXNAELF 508
Db      885 -----GQPGECNTEGSHCVCCQGFSSADGRT-----CEDI 918
QY      509 PEGLRPALPEKH---SSVKESPRYVNLTCSSGKQVP-----GAFGRP-- 547
Db      919 DECVNNTVCDSHGFCDN:TAGSFRCL--CYQGFCAPDGGCGGVNVNECELLSGVCGEAPC 975
QY      548 STPKENFTV-----EPELETNQREVATSCDLSCIVKRTKELRKA:ITLKA----- 595
Db      976 ENVEGSLVCVACADENCEYSFTQCRSRTSTDLVDVQGFKEEKKECYVNLNDAS:CDNV 1035
QY      596 ---VHREOFHLQLS---GXNLDVAKKPRRTSEROAESCGVGCGH-----AEN- 636
Db      1036 LAPNVTKECCCTSGAGWGNCETFP:CPVLGTAEFTEMCPKGGKGVPPAGESSESSEAGSEKY 1095
QY      637 ---QCV-----SCRAGTYVDGARERC-----LCPNGTF 662
Db      1096 KDADECLLFQGE:CKXNGFCLNTRPGVEYCKQGYVDPVKLOCFDMCEQD:SSCIDGQC 1155
QY      663 QNEEGQMTG-----EPCRP:GNS-----S 681
Db      1156 VNTGSYNCFCTHPMVYDASEKRC:RPAESNEQIBETDYYQLCWEHLSDYVYCSRPLVG 1215
QY      682 ALKT-----PEAWNSECGGLCPGSEYSGADGAPAPCOLCALGTFQPEAGRTSCF---- 729
Db      1216 KQTTYTECCCLYGEANGM-QC-ALC-PLKSDSDYALQCLNIPVTCRRQP-YGRDALVDPSE 1271
QY      730 ---PCGGGLATKHQGATSPQD:ETRVQCGSPGHFYNTTTHRC:PCPVG-----TYCPBF 779
Db      1272 QYTPEDAPYV:QDRFLNSFEELQAE-EC--GLJNGCENRCVYVQBGV:CDCLDGVHLDT 1328
QY      780 GKNCVSCPGNTTDFDGTN:ITQCNRRRCGELGCF-----TGVISS--PNY--PNKY 829
Db      1329 AKMTCFD-----VNECELNRRYSLCKNAKINTDGSYKCLCLPGYVSPDKENYCTPLAT 1383
QY      830 PANTE 834
Db      1384 ALNLE 1385

RESULT 10
A38261
masking protein precursor - rat
C:Species: Rattus norvegicus (Norway rat)
C>Date: 14-Jun-1991 #sequence_revision 14-Jun-1991 #text_change 21-Jan-2000
C:Accession: A38261
R:Tsuji, T.; Okada, F.; Yamaguchi, K.; Nakamura, T.
Proc. Natl. Acad. Sci. U.S.A. 87, 8835-8839, 1990
A:Title: Molecular cloning of the large subunit of transforming growth factor type bet
A:Reference number: A38261, MUID:91062373, PMID:2247454
A:Accession: A38261
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1712 <TSU>
A:Cross-references: GB:M55431; NID:Q207285; PIDN:AAA42235.1; PID:Q207286
C:Superfamily: unassigned EGF-related proteins; EGF homology; LDL receptor ligand-bind
F:911-947/Domain: EGF homology <EGF>

Query Match 10.6%; Score 592; DB 2; Length 1712;
Best Local Similarity 24.3%; Pred. No. 4.5e-28;
Matches 230; Conservative 102; Mismatches 305; Indels 310; Gaps 52;
```


A:Accession: A55494

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-1820 <MOR>

A:Cross-references: GB:Z37976

C:Genetics:

A:Gene: GDB:LTBP2

A:Cross-references: GDB:568901

A:Map position: 11pter-11qter

C:Superfamily: unassigned EGF-related proteins; EGF homology

F:1-222-1257/Domain: EGF homology <EGF>

F:1525-1565/Domain: EGF homology <EGF1>

Query Match 10.2%; Score 570; DB 2; Length 1820;

Best Local Similarity 23.2%; Pred. No. 1e-26;

Matches 208; Conservative 110; Mismatches 339; Indels 240; Gaps 45;

QY 46 VDECAQGLDDCHADALCONTPSYKCSKPGVQGEGRQ--CEDIDCEGNEELNGGCVHD-- 101

DB 845 IDRCAGATNVCGPGTCVNLPGYRCVCGSPGYQLHPSOAYCTDNEC---LRDPCQGRGR 901

QY 102 CLNIPGNRYRCTCFDGFMLAHDG--HNCLDVDECLNNGGCO-HTCVNVMGSECCCKECP 158

DB 902 CLNIRVGSFCFCYPTGTTATSGATQECQDINEC-EQPGVCGGQCTNTEGSHCECDGGY 960

QY 159 FL-----SDNQHCTCI-----HRSEGLSCMKND----- 181

DB 961 IMVRKGHCODINECRHPTGCPDGRCVNSPGSYTCLACEEYRGQSG-SCVDVNECLTPGV 1019

QY 182 --HG--CSHICKEAPRGSVACRPGFELAKNORDCILTCNKGNGCQHSDDTAGPBC- 237

DB 1020 CAHKGCTNL-----EGSRFCSCQEQYEYTSDEK-----GCQ-DVDECASTRASCP 1062

QY 238 -----SCHPOYKMHDTGRSLREDTVLEVTESNT--SVVDGSKRVKRGRLM 283

DB 1063 TGLCLNTEGSPACSCAGENCYWNEBGTAC---ED-----L 1094

QY 284 ETCVANNGGCDRTCKDTSTGVHC-SQPVGFTLQDQKTKDIDECQTRNGGC-DHFCRN- 341

DB 1095 DECAFPVGVCPGVCNTAGSPFCKDGGYRSPLDGSDCEDVDECDPQSSCLGSECNT 1154

QY 342 VSGPFGCKKGLLTDEKSCQDVDECSLDRTC--DHSCINHPGTFACNRYG-TLYGF 398

DB 1155 VGSYQCLCPQGFQ-LANGTVCEVNECEMGEBCAPHGECCLASHGSHFFCLCAPGFVSARG 1213

QY 399 THCGDTNECSINNGGCOQCVNTVSGYECQCHPGYKLNHNKDCVVEVXGLLPTSVSPVS 458

DB 1214 TSCQDVDECATTPDPCVGHCVNTEGSEFNCLCEGFPSPSECECVDIDECEDYGDVPCGT 1273

QY 459 LHCGKSGGGDGFRLCHSGIHLSSDVTITRTSVTFKLNKGKSLKNAELFPEGLRPAALPE 518

DB 1274 WKCENSPGSYRCVLGCGPQGFHMAPNGDCI--DIDECANDTMCG----- 1314

QY 519 KH---SSVKESEFYVNLITCSSKQVPGAPGRSTPKEMFIVPEFL-----ETNOKE 567

DB 1315 SHGFCDNTGSRFL---CDQG-----FEISPSGMDCVDVNECE 1350

QY 568 VTASCDLSCIVKTEKRKLARTLRKAVRHQFHLQLSGMNLOVAKXPPR--TSEROAE 625

DB 1351 LMLAVCGAALCENVEGSLCLCASDLBEBYDAGEGTAHGLLEVRVCLRPQDRHAPPTRM 1410

QY 626 SCGVGO-GHA-----ENQVCSKAGTYDQARERCILCPNGTFQNEBQMTCEP 673

DB 1411 DCYSGKGHAPCSSVLGRNTTQAECCCTQGTATGDA---CDLCP-----SEDSAEFSEI 1461

QY 674 CRPGNSGALKTPEAW-----NMSEC-----GGLCQPGE--YSADGFAPCOLCALGTF 719

DB 1462 C--PSKGVIPEGAWTFTQWTYDADECIVFGPLCPNGRCNLNTPGV-C-LCNPQ-F 1516

QY 720 QPEAGRTSCFPGGGLATKHQGATSQDCETRVQCSPGHFYNTT--THRCIRCPVGTIYQPE 778

DB 1517 HYDASHKKC-----EHDCEQD--JACENGECVNEGSEPHFCSPFLTJ--D 1559

QY 779 FGKNNCVSPGNTTTFDGSNTITQCKNRRCGGELGDFDTGYIESPNYPGNYFANITSC 835

DB 1560 LSCQRCWNSTSTEDLPDCHIHMDICWKX-----VTNDVCSEPLRGHRTYTTC 1608

RESULT 13

S78040

fibulin, splice form C precursor - mouse

N:Alternate names: basement-membrane protein BM-90

C:Species: Mus musculus (house mouse)

C:Date: 24-Jul-1998 #sequence revision 24-Jul-1998 #ext_change 02-Aug-2002

C:Accession: S78040; S78560; S36440

R:Pan, T.C.; Kluge, M.; Zhang, R.Z.; Mayer, U.; Timp, R.; Chu, M.L.

Eur. J. Biochem. 215, 733-740, 1993

A:Title: Sequence of extracellular mouse protein BM-90/fibulin and its calcium-depende

A:Reference number: S34968; XUID:93358897; PMID:8354280

A:Accession: S78340

A:Molecule type: mRNA

A:Residues: 1-685 <PAN>

A:Cross-references: EMBL:X70854

R:Pan, T.C.; Kluge, M.; Zhang, R.Z.; Mayer, U.; Timp, R.; Chu, M.L.

submitted to the EMBL Data Library, January 1993

A:Description: Sequence of extracellular mouse protein BM-90/fibulin and its calcium-a

A:Reference number: S36440

A:Accession: S78560

A:Molecule type: mRNA

A:Residues: 1-39, 'P', 41-685 <CHU>

A:Cross-references: EMBL:X70854

C:Genetics:

C:Introns: 568/3

C:Superfamily: fibulin-1; EGF homology

C:Keywords: alternative splicing; basement membrane; calcium binding; extracellular ma

F:1-29/Domain: signal sequence #status predicted <SIG>

F:30-695/Product: fibulin, splice form C #status predicted <MAT>

F:98,537,541/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 10.1%; Score 567; DB 2; Length 685;

Best Local Similarity 27.7%; Pred. No. 5.9e-27;

Matches 159; Conservative 74; Mismatches 159; Indels 182; Gaps 31;

QY 44 EDVDECAQGLDDCHADALCONTPSYKCSKPGVQGEGRQ--CEDIDCEGNEELNGGCVHD-- 98

DB 217 EDINECTGSHNCRGLGESCINTVGSFRQCRDSCGTYELTDNNCKDIDECETGHP-NC 275

QY 99 VHD--C-NIPGNRYC---TCPDGFMLAHDGHNCLDVDECLNNGGCO--HTCVNVMGVS 250

DB 276 PPDFIQTNLGSRFCAPKLCCKSGFTODALG-NCID:NECLCS:SAFCPVQCTCINT:EGSV 334

QY 151 EC-----CKEGEFLSDNQHCTC:HRSEGLSCMKNDH-----GCSHICKEAPRGSA 197

DB 335 TCQKNVNPNGRGY-----HLNEEGTRCVYDDECAFPAPBPCGKHCLNSP-GSFR 383

QY 198 CECRPGFELAKNORDCILTCNKGNGCQHSDDTAGPSCSCHPQYKMHDTGRSCCLERED 257

DB 384 CEKAGYFDFGIRTCV-----DINECQRYP-----GRJC----- 423

QY 258 TVLEVTESNTTSVVDGSKRVKRELLMETCAVNNGGCDRTCKDTSTGVHSCVPGFTQLQD 317

DB 414 -----GKH-----CENTPGSFHCSCSAGFLSVD 437

QY 318 KYTKCKIDSCQTRNGGCDHFCNKIVSGFDGCKKGFLL-TDEKSCQDVDECSL----DRT 373

DB 438 GRSCEDVNEC--LNSPSCQECANVGSYQCYCRRGYQLSDVDGVTCEIDECALPTGGH1 495

QY 374 CDHSC:NHRCPTFACAC-NRGYTLV-QFTHCGDTNEC--SINNGGCOQCVNTVSGYEC-- 427

DB 496 CSYEC:NIPGSFQCCSPSGYRLAPNRCQDIDECVTGIHNCIS:NETCFN:CGSRCLUS 555

QY 428 -CQHPGVK-----JHNKKKOCVEVKGLL-----PVSLSHCKGSGGG 467

DB 556 FECPENYRSADTRCARLPCHEN-GSCPRLPLRITYHLSPFNIQVPAVFMGSSAV 614

QY 468 DCGFLKCHSGIHLSSDVTITRTSVTFKLNKGKSLKNAELFPEGLRPAALPEKXSYKESF 527

```

Db      615 PG-----DSMQLAITAGNEGFFTRKV-----SHRSGV----- 643
QY      528 RYVNLTCSSGKQVGPAGRPSTPKEMFIVTEFEL 561
Db      644 --VALT---KPIP-----EPRDLTLTKVMDL 664

RESULT 14
C36346
Fibulin-1 precursor, splice form C - human
N:Alternate names: fibulin C
N:Contains: fibulin-1 splice form A; fibulin-1 splice form C
C:Species: Homo sapiens (man)
C:Date: 19-Apr-1991 #sequence_revision 19-Apr-1991 #text_change 02-Aug-2002
C:Accession: C36346; A36346; A32826
R:Argaves, W.S.; Tran, H.; Burgess, W.H.; Dickerson, K.
J. Cell Biol. 111, 3155-3164, 1990
A:Title: Fibulin is an extracellular matrix and plasma glycoprotein with repeated domain
A:Reference number: A36346; MUID:91100426; PMID:2265669
A:Accession: C36346
A:Molecule type: mRNA
A:Residues: 1-683 <ARG>
A:Cross-references: GB:X53743; NID:g31418; PIDN:CAA37772.1; PID:g31419
A:Accession: A36346
A:Molecule type: mRNA
A:Residues: 1-566 <AR2>
A:Cross-references: GB:X53741; NID:g31414; PIDN:CAA37770.1; PID:g31415
R:Argaves, W.S.; Dickerson, K.; Burgess, W.H.; Ruoslahti, E.
Cell 58, 621-629, 1989
A:Title: Fibulin, a novel protein that interacts with the fibronectin receptor beta-sub
A:Reference number: A32826; MUID:89354537; PMID:2527614
A:Accession: A32826
A:Molecule type: protein
A:Residues: 30-35, SX, 38-40, 'SH', 43-44 <AR3>
C:Genetics:
A:Gene: GDB:FBLN1; FBLN
A:Cross-references: GDB:278285; OMIM:135820
A:Map position: 22q13.3-22q13.3
C:Superfamily: fibulin-1; EGF homology
C:Keywords: alternative splicing; glycoprotein
F:1-29/Domain: signal sequence #status predicted <SIG>
F:30-683/Product: fibulin-1 splice form C #status predicted <MAT>
F:180-214/Domain: EGF homology <EGF>
F:485-523/Domain: EGF homology <EGF>
F:98,535,539/Binding site: carbohydrate (Asn) (covalent: #status predicted

Query Match      10.1%; Score 565; DB 2; Length 683;
Best Local Similarity 27.9%; Pred. No. 7.6e-27;
Matches 160; Conservative 77; Mismatches 155; Indels 192; Gaps 31;

QY      44 EDVDSACGLDDCHADALCONFTSYKC-----SKPGYQ-GEGRQCEIDECGHELDGCG 98
Db      215 EDVNEGIGTGHSCRLGSCINTVSGFRQDRSSCGTGHELTEDNSCKDIDECESGIR-NC 273
QY      99 VHD--CLNIPGNRYC-----TCDFGFMALAHGNCILVDDECLENNCGCQ--HTCVNVMGSS 150
Db      274 LPDFICQNTLGSFRCPKPLQCKSGFIQDALG-NCIDINECLISAFPCFIGHTC-NTGESSY 332
QY      151 EC-----CKKEGFFLSDNQHTCIHRSEGLSMKNKH-----GCSHICKKAPRGSA 197
Db      333 TCOKNVNPGCGY-----HLNEGTRCVVDDECAPPAPGCGKGRWCNVP-GSPR 391
QY      198 CECRPGFELAKNQRCILTCNHNNGGCGHSCDDTAJGPECSCHPQYKXHTDGRSCLERD 257
Db      382 CEKGTGYFDGISRCMV-----DVNECQRYP-----GRLC----- 411
QY      258 TVLEVTESNTTSVPDGRKVRKRLMETCAVNNCGCDKTCCKDTSTGVHSCSPGVGFTQLD 317
Db      * 412 -----GKK-----CENTLGSYLSCSVGFR-LSD 435
QY      318 GXTCKDIDECQTRNGCDHFCKNIIVGSPDCKGCKGFKLL-TDEKSCQVDKCSL---DRT 373

```

```

Db      436 GRSCSDINECS--SSPCSCQCANVVGSYCCYCRGYQLSDVDGVTCEIDECALPTGGHI 493
QY      374 CMHSCINHPGTACAC-NRGYTLY-GPTHGDTNCC--SINNGGQQCVQVNTVVGSEYEC-- 427
Db      494 CSYRCINIFGSPQSCPPSSGYRLAPNGRNCQDIDCVTGIHNCS-NETCFKIQGAFRCJA 553
QY      428 -CHPGYK-----LHWNKKDCVEVKLL-----PTSV-SPEVSLHCKCKSGG 467
Db      554 FPCPENYRRAATRCERLPCHEN-RECSKLPAPRIYYHLSPFTNQAFAVVRXGPPSSAV 612
QY      468 DGCFLRCHSGIHLSEVITIRISVTFKLNKCSLKNALFPEGUPPALPEKHSSVKSEF 527
Db      613 PG-----DSMQLAITGNEGFFTRKV-----SPRSGV----- 641
QY      528 RYVNLTCSSGKQVGPAGRPSTPKEMFIVTEFEL 561
Db      642 --VALT---KPIP-----EPRDLTLTKVMDL 662

RESULT 15
A55184
Fibulin-2 precursor - human
N:Alternate names: protein DKFP586A1519.1
C:Species: Homo sapiens (man)
C:Date: 27-Jan-1995 #sequence_revision 27-Jan-1995 #text_change 08-Sep-2002
C:Accession: A55184; T08744
R:Zhang, R.Z.; Ban, T.C.; Zhang, Z.Y.; Mattei, M.G.; Timp, R.; Chu, M.L.
Genomics 22, 425-430, 1994
A:Title: Fibulin-2 (FBLN2): human cDNA sequence, mRNA expression, and mapping of the g
A:Reference number: A55184; MUID:95-04855; PMID:7806230
A:Accession: A55184
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1184 <ZHA>
A:Cross-references: GB:X82494; NID:g575232; PIDN:CAA57876.1; PID:g575233
R:Mambutti, R.; Heubner, D.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, March 1999
A:Reference number: Z16471
A:Accession: T08744
A:Molecule type: mRNA
A:Residues: 656-719; QDECLMGADHCSRROFCVNTLGSFYCVNHTVLCADGY-LNAHRKCVQ, 720-853, T, 855-1
A:Cross-references: EMBL:AJ050035
A:Experimental source: adult uterus; clone DKFP586A1519
C:Genetics:
A:Gene: GDB:FBLN2
A:Cross-references: GDB:293037; OMIM:135821
A:Map position: 3p25-3p24
A:Superfamily: fibulin-2; EGF homology; von Willebrand factor type C repeat homology
C:Keywords: alternative splicing; extracellular matrix
F:1-27/Domain: signal sequence #status predicted <SIG>
F:28-1184/Product: fibulin-2 protein #status predicted <MAT>
F:955-941/Domain: EGF homology <EGF>

Query Match      10.0%; Score 560; DB 2; Length 1184;
Best Local Similarity 29.1%; Pred. No. 2.7e-26;
Matches 141; Conservative 68; Mismatches 155; Indels 138; Gaps 25;

QY      44 EDVDECAAGLDCHADALCONFTSYKC-----SKPGYQGEGRQCEIDECG--NELNG 96
Db      718 EDNECVTLTHCSRGHCYNTLGSFHCYKALTCPGVALKDGECEDECAAGTHTCGP 777
QY      97 GCVHDCNLIPGNRYC-----TCDFGFMALAHGNCILVDDECLENNCGCQ--HTCVNVMGSS 150
Db      778 GFL--QNTKGSFYCOARQFCYGLQDPBG-NCVDINECTLSLSEPCRPGFSCINTVGSY 834
QY      151 EC-----CKKEGFFLSDNQHTCIHRSEGLSMKNKHGCG--SHICKKAPRGSA 203
Db      835 TCORNPICARGYHASDDGAKCVNNE-----CETGVHRCGEGQVCHNLP-GSYRCDCKAG 889
QY      204 FSLAKNQRCILTCNHNNGGCGHSCDDTAGDEPSCCHFCYKXHTDGRSCLERD 263
Db      890 F-----ORDAF-----GRGCLD----- 901

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 22, 2003, 11:17:23 ; Search time 26 Seconds
(without alignments)
1806.911 Million cell updates/sec

Title: US-09-747-371-2

Perfect score: 5605

Sequence: 1 MGAGNRNPGAAWVLLLL.....PRSFIRLLRSKVSFRPRPX 999

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	727	13.0	2911	1 FBN2_HUMAN	P35556 homo sapien
2	724	12.9	2907	1 FBN2_MOUSE	Q61555 mus musculus
3	692	12.3	956	1 MTN2_MOUSE	O08746 mus musculus
4	691	12.3	2871	1 FBNI_BOVIN	P98133 bos taurus
5	691	12.3	2871	1 FBNI_HUMAN	P35555 homo sapien
6	690	12.3	2871	1 FBNI_MOUSE	O61554 mus musculus
7	684.5	12.2	956	1 MTN2_HUMAN	O00339 homo sapien
8	678	12.1	2871	1 FBNI_PIG	Q95V36 sus scrofa
9	624.5	11.1	1394	1 LTBS_HUMAN	P22064 homo sapien
10	624.5	11.1	1595	1 LTBL_HUMAN	Q14766 homo sapien
11	601	10.7	1389	1 LTBL_MOUSE	Q8CG18 mus musculus
12	601	10.7	1713	1 LTBL_MOUSE	Q8CG19 mus musculus
13	592	10.6	1121	1 LTBI_RAT	Q00918 rattus norv
14	577.5	10.3	1221	1 FB2_MOUSE	P37889 mus musculus
15	567	10.1	703	1 FB1_HUMAN	P23142 homo sapien
16	562	10.0	704	1 FB1_CHICK	Q73775 gallus gal
17	560	10.0	1184	1 FB2_HUMAN	P98095 homo sapien
18	558	10.0	705	1 FB1_MOUSE	O08879 mus musculus
19	529.5	9.4	798	1 FB1_CAEEL	O77469 caenorhabdi
20	526.5	9.4	2321	1 NTC3_HUMAN	Q9UM47 homo sapien
21	524.5	9.4	598	1 FB1_CERAE	Q8MJ99 cercopithe
22	524.5	9.4	2318	1 NTC3_MOUSE	Q61982 mus musculus
23	516.5	9.2	2319	1 NTC3_RAT	Q9X172 rattus norv
24	509.5	9.1	2703	1 NOTC_DROME	P07207 drosophila
25	501.5	8.9	2524	1 NOTC_XENLA	P21783 xenopus lae
26	500.5	8.9	681	1 FB1_BRARE	Q42182 brachydanio
27	500	8.9	2470	1 NTC2_MOUSE	C35516 mus musculus
28	489.5	8.7	1964	1 NTC4_MOUSE	P31695 mus musculus
29	489	8.7	2531	1 NTC1_MOUSE	Q01705 mus musculus
30	489	8.7	2531	1 NTC1_RAT	Q07008 rattus norv
31	485	8.7	2471	1 NTC2_HUMAN	Q04721 homo sapien
32	484.5	8.6	2556	1 NTC1_HUMAN	P46531 homo sapien
33	484	8.6	2437	1 NTC1_BRARE	P46530 brachydanio

34	479	8.5	2471	1 NTC2_RAT	Q9QW30 rattus norv
35	467	8.3	2003	1 NTC4_HUMAN	Q99466 homo sapien
36	466	8.3	1064	1 FB1_STRPU	P10079 strongyloce
37	464.5	8.3	3767	1 MUA3_CAEEL	P34576 caenorhabdi
38	443.5	7.9	1429	1 L112_CAEEL	P14595 caenorhabdi
39	438	7.8	1217	1 EGF_MOUSE	P01132 mus musculus
40	434	7.7	448	1 FB15_RAT	Q9WV08 rattus norv
41	431.5	7.7	1133	1 EGF_RAT	P07522 rattus norv
42	429	7.7	1238	1 JAG2_HUMAN	Q9Y219 homo sapien
43	426.5	7.6	448	1 FB15_HUMAN	Q9UBX5 homo sapien
44	426.5	7.6	448	1 FB15_MOUSE	Q9WV09 mus musculus
45	421.5	7.5	1218	1 JAG1_HUMAN	P78504 homo sapien

ALIGNMENTS

RESULT 1	
FBN2_HUMAN	
AC FBN2_HUMAN	STANDARD; PRT: 2911 AA.
AC P35556;	
DT 01-JUN-1994 (Rel. 29, Created)	
DT 01-FEB-1996 (Rel. 33, Last sequence update)	
DT 28-FEB-2003 (Rel. 41, Last annotation update)	
DE Fibrillin 2 precursor.	
GN FBN2.	
OS Homo sapiens (Human).	
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
OX NCBI_TaxID=9606;	
RN 1.	
RP SEQUENCE FROM N.A.	
RX MEDLINE=94185150; PubMed=8120105;	
RA Zhang H., Apfelroth S.D., Hu W., Davis E.C., Sarguineti C.,	
RA Bonadio J., Yecham R.P., Ramirez F.,	
RT "Structure and expression of fibrillin-2, a novel microfibrillar	
RT component preferentially located in elastic matrices.";	
RL J. Cell Biol. 124:855-863(1994).	
RN (2)	
RP SEQUENCE OF 752-1505 FROM N.A.	
RX MEDLINE=91304567; PubMed=1552206;	
RA Lee B., Godfrey M., Vitale E., Hori H., Mattei M.-G., Sarfarazi M.,	
RA Tsipouras P., Ramirez F., Hollister D.;	
RT "Linkage of Marfan syndrome and a phenotypically related disorder to	
RT two different fibrillin genes.";	
RL Nature 352:330-334(1991).	
RN (3)	
RP VARIANTS CCA TYR-1252 AND SER-1433, AND VARIANT ILE-964.	
RX MEDLINE=96083599; PubMed=7493032;	
RA Putnam E.A., Zhang H., Ramirez F., Milewicz D.M.;	
RT "Fibrillin-2 (FBN2) mutations result in the Marfan-like disorder,	
RT congenital contractural arachnodactyly.";	
RL Nat. Genet. 11:456-458(1995).	
RN (4)	
RP VARIANTS CCA HIS-1114.	
RX MEDLINE=98407789; PubMed=9737771;	
RA Babcock D., Gasner C., Francke U., Maslen C.;	
RT "A single mutation that results in an asp-to-his substitution and	
RT partial exon skipping in a family with congenital contractural	
RT arachnodactyly.";	
RL Hum. Genet. 103:22-28(1998).	
RN (5)	
RP VARIANTS CCA PHE-1141 AND TRP-1252.	
RX MEDLINE=20259236; PubMed=0797416;	
RA Beilen S., Zhou G., Wang M., Der Kaloustian V.M., Pagon R.A.,	
RT Godfrey M.;	
RT "Two novel fibrillin-2 mutations in congenital contractural	
RT arachnodactyly.";	
RL Am. J. Med. Genet. 92:7-12(2000).	
CC -!- FUNCTION: STRUCTURAL COMPONENT OF CONNECTIVE TISSUE MICROFIBRILS	
CC THAT BINDS CALCIUM. FIBRILLIN-2-CONTAINING MICROFIBRILS REGULATE	
CC THE EARLY PROCESS OF ELASTIC FIBER ASSEMBLY.	
CC -!- DISEASE: DEFECTS IN FBN2 ARE THE CAUSE OF CONGENITAL CONTRACTURAL	

Query Match 13.0%; Score 727; DB 1; Length 2911;
Best Local Similarity 26.5%; Pred. No. 2e-40;
Matches 255; Conservative 94; Mismatches 301; Indels 312; Gaps 51;

QY 45 DVDECAAGGDDCHADALCONTTSYKSCCKPQYOG--EGRCQEDIDECGNEIAGCVHDC 102
D 1199 DINECSLS-DNLCRNKGVNMTGTQVCSNPGVQATPDQGGTDIDEC-MIMWGGCDTQC 1256
QY 103 LNIPIGYRCTCDGFMALHGHNCIDVDRECLNN-----GGQHTCVNVMGSEYCCCKEG 157
D 1257 TNSGSEYSCSEGVYALPDGRSCADIDEC-ENNPEDICGG--OCTNIPGEYRCLCYDG 1312
QY 158 FFLSNQHTCIHRSEBGLSCMKDHGCSHICK-----EAPRGVACECPRPGLFLANKQDC 213
D 1313 FNASMDMKTCIDVNECDLN-----SNICMFGECENTKGSFICHCQLGYSVKGGTTC 1354
QY 214 --LLTCHNGGC--QHSDDTADGPECSCHPQYXMHDTGRSLERETVLEVTESN-TS 269
D 1365 TDVDECEIHAHNCMDHASCLNIPGSKSCREGW--IGNGIKIDBLDEC-----SNGT- 1415
QY 270 VVDGKRVRRLLMETCAVNNGGCDRTCKDTSTGVHSCCPVGTTLQLDGKTKCKDIDECCT 329
D 1416 -----HCCSIN-----AQCNTPGSYRCACSEGT--GGGTCSVDDECAE 1454
QY 330 RNGGCDH-FCKNIVGFDGCKKGFLLTDEKSCQDVDECSLDRTC-DHSCINHPGTAC 387
D 1455 NINLCENGQCLNVPYAYRCEBNGFTPASDRSRSCDIDECSPQNICVSTCNLNPVFFHC 1514
QY 388 ACNRGYTL-YGFTHCGTNECS--IN--NGGCOQCVNVTGSEYCOCHPYKHLWNKKDC 442
D 1515 ICDDGYELDTGNCNTDIDECADPNCVNG---LCVNTPGRYECNCFDFOLNPTGVGC 1570
QY 443 VEYK-GLLPTSVPRLVHCGSGGDDGFLRCHSGIHLSSDVTIRTSVTFKLNEGRCS 501
D 1571 VDNRVGNCYLKGFPR-----GDGS-LSGNTFI---GVGYSRSCC-----CS 1608
QY 502 LKNAELPPGLPALPEKHSVKESPRYNLTCSGKQVGPAGRPSTPKENFTVSEFEL 561
D 1609 LKAWGNPCETCPV-----NSTEYVTL-----CFGGEGRPNP-----IILLED 1649
QY 562 ETNQKEVTASCD-LSGIVKRTKRLKA-RTLKAVHREQFHLQSGMKV-DVAKKFPRTS 620
D 1650 IDECQELPLGCOGGNCI-----N 1667
QY 621 EROAESCGVGGH--AE-----NOCVS-----CRAGTYDGASERCILCP----- 658
D 1668 TFGSFQCECPQYLSYEDTRICEDIFAHGVCVCGEGTCYNTLGNVTCICPEYVQVNG 1727
QY 659 -----NG-TFQNE-----EGQNTCEPCRPQNSG-- 681
D 1728 GHNCMDMRKSFYRSYNGTTCENELPFNVTKRMCOCTYVNGKAGNKPCEPCPTGTADFK 1787
QY 682 -----ALKTPAEMWMBEC----- 694
D 1788 TICGNTPGFTDILHTGXAVDIDECKEIPGICANGVCINQISFRCEPTGFSYNDLLVC 1847
QY 695 -----GGLCQ-----PGEYSADGPAFCQLCALG-TFQPEAG---RTSCFP--- 730
D 1848 EDIDECNGDNLCQRNADCINSFGRCE-----CAAGFKLSPNACVDSNECLEIFN 1900
QY 731 -CGGLATKHQATSFQ-DETRVQCSPGHFYNTTTHRCIR--CPVGTVPQEPGKNCVUS 786
D 1901 VCSHGLCVDLQG--SYQICHNHGFKASQDOTCMQVDECEHFCGNGTKNTVGSYNCLC 1958
QY 787 CPG-----NTTDFDPSGTNI--TCKNRRCCGELGDF-----TGYESPNYPGVNPANT 833
D 1959 YPQFELTHNDCDIDECSSFFQVCVRNCRGCFNEIGSKFLCONEGYELTPDSKNCIDN- 2017
QY 834 EC 835
D 2018 EC 2019

RESULT 2
FN2 MOUSE STANDARD; PRT; 2907 AA.
AC Q61555; C61957;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Fibrillin 2 precursor.
GN FN2 OR FN2-2.
CS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
PN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=95263670; PubMed=7744963;
RA Zhang H., Hu W., Ramirez F.;
RT "Developmental expression of fibrillin genes suggests heterogeneity
RT of extracellular microfibrils.";
RJ J. Cell Biol. 129:1165-1176(1995).
RN [2]
RP SEQUENCE OF 210-317 P50Y N.A.
RX MEDLINE=94140368; PubMed=637578;
RA Li X., Pereira L., Zhang H., Sanguinetti C., Ramirez F., Bonadio J.,
RA Francke U.;
RT "Fibrillin genes map to regions of conserved mouse/human synteny on
RT mouse chromosomes 2 and 18.";
RL Genomics 18:667-672(1993).
CC -- FUNCTION: STRUCTURAL COMPONENT OF CONNECTIVE TISSUE MICROFIBRILS
CC THAT BINDS CALCIUM. FIBRILLIN-1-CONTAINING MICROFIBRILS PROVIDE
CC LONG-TERM FORCE BEARING STRUCTURAL SUPPORT.
CC -- SIMILARITY: Contains 47 EGF-like domains.
CC -- SIMILARITY: Contains 7 TGF-beta binding protein (TGFBP) domains.
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
DR EMBL; L3979C; AAA74908.1; -;
DR EMBL; S69359; AAC60685.1; -;
DR PIR; A57278; A57278.
DR HSSP; P35555; 1EMN.
DR MGJ; MGJ19549C; Fbr2.
DR GO; GO:0030326; P, limb morphogenesis; IMP.
DR InterPro; IPR000152; Asx hydroxyl.
DR InterPro; IPR001861; EGF-Ca.
DR InterPro; IPR001438; EGF-1.
DR InterPro; IPR006209; EGF-like.
DR InterPro; IPR002212; Fibrin-assoc.
DR Pfam; PF00308; EGF; 45.
DR Pfam; PF00683; TB; 9.
DR PRINTS; PR00010; EGFBL0CD.
DR SMART; SM00179; EGF CA; 43.
DR PROSITE; PS00010; ASX HYDROXYL; 43.
DR PROSITE; PS00022; EGF_1; 2.
DR PROSITE; PS01186; EGF_2; 36.
DR PROSITE; PS01187; EGF CA; 43.
DR Extracellular matrix; Calcium-binding; Glycoprotein; EGF-like domain;
DR Repeat; Signal; Multigene family.
FT SIGNAL 28
FT CHAIN 29 2907
FT FIBRILLIN 2.
FT DOMAIN 112 142
FT EGF-LIKE 1.
FT DOMAIN 145 176
FT EGF-LIKE 2.
FT DOMAIN 176 208
FT EGF-LIKE 3.
FT DOMAIN 276 317
FT EGF-LIKE 4.
FT DOMAIN 318 359
FT EGF-LIKE 5.
FT REPEAT 360 426
FT TGFBP 1.
FT DOMAIN 487 527
FT EGF-LIKE 6.

```
FT DOMAIN 528 567 EGF-LIKE 7, CALCIUM-BINDING.
FT DOMAIN 568 609 EGF-LIKE 8, CALCIUM-BINDING.
FT DOMAIN 610 650 EGF-LIKE 9, CALCIUM-BINDING.
FT DOMAIN 651 691 EGF-LIKE 10, CALCIUM-BINDING.
FT REPEAT 692 760 TGFBP 2.
FT DOMAIN 761 802 EGF-LIKE 11, CALCIUM-BINDING.
FT DOMAIN 803 844 EGF-LIKE 12, CALCIUM-BINDING.
FT DOMAIN 845 883 EGF-LIKE 13, CALCIUM-BINDING.
FT DOMAIN 948 989 EGF-LIKE 14, CALCIUM-BINDING.
FT REPEAT 990 1065 TGFBP 3.
FT DOMAIN 1066 1107 EGF-LIKE 15, CALCIUM-BINDING.
FT DOMAIN 1108 1150 EGF-LIKE 16, CALCIUM-BINDING.
FT DOMAIN 1151 1192 EGF-LIKE 17, CALCIUM-BINDING.
FT DOMAIN 1193 1234 EGF-LIKE 18, CALCIUM-BINDING.
FT DOMAIN 1235 1275 EGF-LIKE 19, CALCIUM-BINDING.
FT DOMAIN 1276 1317 EGF-LIKE 20, CALCIUM-BINDING.
FT DOMAIN 1318 1359 EGF-LIKE 21, CALCIUM-BINDING.
FT DOMAIN 1360 1400 EGF-LIKE 22, CALCIUM-BINDING.
FT DOMAIN 1401 1441 EGF-LIKE 23, CALCIUM-BINDING.
FT DOMAIN 1442 1483 EGF-LIKE 24, CALCIUM-BINDING.
FT DOMAIN 1484 1524 EGF-LIKE 25, CALCIUM-BINDING.
FT DOMAIN 1525 1565 EGF-LIKE 26, CALCIUM-BINDING.
FT REPEAT 1566 1642 TGFBP 4.
FT DOMAIN 1643 1684 EGF-LIKE 27, CALCIUM-BINDING.
FT DOMAIN 1685 1726 EGF-LIKE 28, CALCIUM-BINDING.
FT REPEAT 1727 1800 TGFBP 5.
FT DOMAIN 1801 1842 EGF-LIKE 29, CALCIUM-BINDING.
FT DOMAIN 1843 1884 EGF-LIKE 30, CALCIUM-BINDING.
FT DOMAIN 1885 1926 EGF-LIKE 31, CALCIUM-BINDING.
FT DOMAIN 1927 1965 EGF-LIKE 32, CALCIUM-BINDING.
FT DOMAIN 1966 2008 EGF-LIKE 33, CALCIUM-BINDING.
FT DOMAIN 2009 2048 EGF-LIKE 34, CALCIUM-BINDING.
FT DOMAIN 2049 2090 EGF-LIKE 35, CALCIUM-BINDING.
FT REPEAT 2091 2163 TGFBP 6.
FT DOMAIN 2164 2205 EGF-LIKE 36, CALCIUM-BINDING.
FT DOMAIN 2206 2245 EGF-LIKE 37, CALCIUM-BINDING.
FT DOMAIN 2246 2286 EGF-LIKE 38, CALCIUM-BINDING.
FT DOMAIN 2287 2330 EGF-LIKE 39, CALCIUM-BINDING.
FT DOMAIN 2331 2372 EGF-LIKE 40, CALCIUM-BINDING.
FT REPEAT 2373 2441 TGFBP 7.
FT DOMAIN 2442 2483 EGF-LIKE 41, CALCIUM-BINDING.
FT DOMAIN 2484 2524 EGF-LIKE 42, CALCIUM-BINDING.
FT DOMAIN 2525 2563 EGF-LIKE 43, CALCIUM-BINDING.
FT DOMAIN 2564 2606 EGF-LIKE 44, CALCIUM-BINDING.
FT DOMAIN 2607 2646 EGF-LIKE 45, CALCIUM-BINDING.
FT DOMAIN 2647 2687 EGF-LIKE 46, CALCIUM-BINDING.
FT DOMAIN 2688 2727 EGF-LIKE 47, CALCIUM-BINDING.
FT DISULFID 115 124 BY SIMILARITY.
FT DISULFID 119 130 BY SIMILARITY.
FT DISULFID 132 141 BY SIMILARITY.
FT DISULFID 149 159 BY SIMILARITY.
FT DISULFID 153 164 BY SIMILARITY.
FT DISULFID 166 175 BY SIMILARITY.
FT DISULFID 180 190 BY SIMILARITY.
FT DISULFID 184 196 BY SIMILARITY.
FT DISULFID 198 207 BY SIMILARITY.
FT DISULFID 280 292 BY SIMILARITY.
FT DISULFID 287 301 BY SIMILARITY.
FT DISULFID 303 316 BY SIMILARITY.
FT DISULFID 322 334 BY SIMILARITY.
FT DISULFID 329 343 BY SIMILARITY.
FT DISULFID 345 358 BY SIMILARITY.
FT DISULFID 498 512 BY SIMILARITY.
FT DISULFID 514 526 BY SIMILARITY.
FT DISULFID 532 542 BY SIMILARITY.
FT DISULFID 537 551 BY SIMILARITY.
FT DISULFID 553 566 BY SIMILARITY.
FT DISULFID 572 584 BY SIMILARITY.
FT DISULFID 579 593 BY SIMILARITY.
FT DISULFID 595 608 BY SIMILARITY.
FT DISULFID 614 625 BY SIMILARITY.
FT DISULFID 620 634 BY SIMILARITY.
```

```
FT DISULFID 636 642 BY SIMILARITY.
FT DISULFID 655 666 BY SIMILARITY.
FT DISULFID 661 675 BY SIMILARITY.
FT DISULFID 677 690 BY SIMILARITY.
FT DISULFID 765 777 BY SIMILARITY.
FT DISULFID 772 786 BY SIMILARITY.
FT DISULFID 788 801 BY SIMILARITY.
FT DISULFID 807 819 BY SIMILARITY.
FT DISULFID 824 828 BY SIMILARITY.
FT DISULFID 830 843 BY SIMILARITY.
FT DISULFID 849 859 BY SIMILARITY.
FT DISULFID 854 868 BY SIMILARITY.
FT DISULFID 870 883 BY SIMILARITY.
FT DISULFID 952 964 BY SIMILARITY.
FT DISULFID 959 973 BY SIMILARITY.
FT DISULFID 975 988 BY SIMILARITY.
FT DISULFID 1070 1082 BY SIMILARITY.
FT DISULFID 1077 1091 BY SIMILARITY.
FT DISULFID 1093 1106 BY SIMILARITY.
FT DISULFID 1112 1124 BY SIMILARITY.
FT DISULFID 1119 1133 BY SIMILARITY.
FT DISULFID 1135 1149 BY SIMILARITY.
FT DISULFID 1155 1167 BY SIMILARITY.
FT DISULFID 1162 1176 BY SIMILARITY.
FT DISULFID 1178 1191 BY SIMILARITY.
FT DISULFID 1197 1209 BY SIMILARITY.
FT DISULFID 1204 1218 BY SIMILARITY.
FT DISULFID 1220 1233 BY SIMILARITY.
FT DISULFID 1239 1250 BY SIMILARITY.
FT DISULFID 1246 1259 BY SIMILARITY.
FT DISULFID 1261 1274 BY SIMILARITY.
FT DISULFID 1280 1292 BY SIMILARITY.
FT DISULFID 1287 1301 BY SIMILARITY.
FT DISULFID 1303 1316 BY SIMILARITY.
FT DISULFID 1322 1334 BY SIMILARITY.
FT DISULFID 1329 1343 BY SIMILARITY.
FT DISULFID 1345 1358 BY SIMILARITY.
FT DISULFID 1364 1377 BY SIMILARITY.
FT DISULFID 1371 1386 BY SIMILARITY.
FT DISULFID 1388 1399 BY SIMILARITY.
FT DISULFID 1405 1418 BY SIMILARITY.
FT DISULFID 1412 1427 BY SIMILARITY.
FT DISULFID 1429 1440 BY SIMILARITY.
FT DISULFID 1446 1458 BY SIMILARITY.
FT DISULFID 1453 1467 BY SIMILARITY.
FT DISULFID 1463 1482 BY SIMILARITY.
FT DISULFID 1469 1499 BY SIMILARITY.
FT DISULFID 1494 1508 BY SIMILARITY.
FT DISULFID 1510 1523 BY SIMILARITY.
FT DISULFID 1529 1542 BY SIMILARITY.
FT DISULFID 1535 1549 BY SIMILARITY.
FT DISULFID 1551 1564 BY SIMILARITY.
FT DISULFID 1647 1653 BY SIMILARITY.
FT DISULFID 1654 1669 BY SIMILARITY.
FT DISULFID 1670 1683 BY SIMILARITY.
FT DISULFID 1689 1701 BY SIMILARITY.
FT DISULFID 1696 1710 BY SIMILARITY.
```

Query Match 12.3%; Score 724; DB 1; Length 2907;

Best Local Similarity 26.1%; Pred. No. 3,2e-40;

Matches 25; Conservative 98; Xismatches 300; Indels 314; Gaps 51;

Qy 45 DVDECAQGLDCHADALCQNTFTSYKSCCKPGYQG--EGROCEDIDEGONEJNGCVDHC 102

Db 1193 DINECSLS-DNLCRNGKCVNMGTYQCSNCPGYQATPDQGGCTDIDEC-MINWGSCDTC 1250

Qy 103 LNIIPGNYRTCTPDGFMALHDGHNCILDVDECLENN-----GGCQHTCVNMGSEYECCKEG 157

Db 1251 TNSGSEYECSCSEGVALMPDGRSCADIDEC-ENNPDI CDGG---OCTNIPGEYRCLCYDG 1306

Qy 158 FFLSDNCHTCIHRSEEGLSYKNDHCCHICK-----EAPRGSVACECRPGFELAKNORDC 213

Db 1307 FMASMDNKTCDVNECCLN-----FNICMFGECENTKGSFICHQGLGVYKKGTCGC 1358

```

QY 214 --ILTNHNGGCG--CHSCDDTADGPBSCCHPOYXHTDGRSCLREDTVLEVTESNTTS 269
Db 1359 TDVDECEIQAHCMDHASCINVPFGSKSCREGWV--GNGIKCIDLDEC-----ANGT- 1403
QY 270 VVDGKRVKRLMETCAVNNGGCDRTCKDSTGVHSCSPVGFQLQDGKCKDIDECOT 329
Db 1410 -----HOCNIN-----AQCNTPGSVRCACSEGT--GGGFTCSVDVECAE 1448
QY 330 RGGGCDH-PCKNIVSFDGCKKPKLLTDXSCQDVDECSLDRCTCDH-SCINHPGTAC 387
Db 1449 NTNLCENGQCLNVPAYR-CECEWGTTPASDSRSCODIDECSPQNCVFGTCNNLPGMPC 1508
QY 388 ANRGHTL-YGTHGCDTNECS--IN--NGGQQCVNVTVSQYQCQHPGKHLNKKQ 442
Db 1509 ICDDGYELDRGTGCTDIDECADPNVNGV-----LCVNTPGRYECNCPDFQLNATGVGC 1564
QY 443 VEVK-GLLPTSVPVSLHCGSGGSGGCFLRCHSGIHLSDVTTIRTSVTFKLNKCS 501
Db 1565 VDNRVGNCVLFKGP--GDS-LSGNT-----EAGVGVSRSQC-----CS 1602
QY 502 LKNABLF-EGURPALPERKSHVKESFRVNLTCSSGKQVPGAPGRSTPKEMFTTVEFE 560
Db 1603 LCKANGNCPCTPPVNSTEYHL-----CPGEGFRNP-----ITILE 1642
QY 561 LETNOKEVTASCD-LSCIVRTEKRLKAIKRLKAVHREQLHLSGNLDVAKKPPRT 619
Db 1643 DIDECELPGLCGGNCI-----NOCVS-----CRAGTYDGAERCIICLP----- 658
QY 620 SERAESCGVGQGH--AE-----NOCVS-----CRAGTYDGAERCIICLP----- 658
Db 1661 NTFSGFCBCPOQYLYSEBTRICEDIDECFAHPGVGGPGCTVNTLGNVTCICPPYKQVN 1720
QY 659 -----NG-TFQNE-----EQMTCPCPCPRGNSG- 661
Db 1721 GGHNCMDMRKSCFYRSYNGTTCENELPFNVTKRMCCCTVYVCKAGNKCEPCPTGTADP 1782
QY 682 -----ALKTPEAWNSEC----- 694
Db 1781 KTCIGNIPGFTDIHTGKAVDIDEKEIPGICANGVCINQIGSFCECPTGFSYNDJLLV 184C
QY 695 -----GGIQC-----PGYSADGPAPCOLCALG-TFOEAG---RTSCFP-- 730
Db 1841 CEDIDECNSGNDLQCRNACINSPGSRCE-----CAAGFKLSPGACVDRNCLCZP 1893
QY 731 --CGGLATKHGATSFO-DCETRVCSPGHFYNTTTHRCIR--CPVGTYPQEFKNNCV 785
Db 1894 NVCSHGLCVDLQ--SYQICNNGPKASQDQTMCDVDECHPCANGTCKNTVGSYNCL 1951
QY 786 SCFG-----NTTDFDGSNTI--TCKNRRCGGELGDF-----TGYTESPNYPCNYFAN 832
Db 1952 CYPGFELTHNDCLDIDECSPFGQVCRNGRCFNEIGSKFCJCNEGYELTPDGKNCIDTN 2011
QY 833 TEC 835
Db 2012 -EC 2013

RESULT 3
MTN2 MOUSE STANDARD; PRT; 956 AA.
AC C08716;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Matrilin-2 precursor.
GN MATN2
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
RN NCBI_TaxID=10090;
RP [1]
SEQUENCE FROM N.A.

```

```

EC STRAIN=BALE/c; TISSUE=Limb;
RX MEDLINE=97238943; PubMed=9083061;
RA Deak F., Piecha D., Bachrati C., Paulsson M., Kiss I.;
RT "Primary structure and expression of matrilin-2, the closest relative
RT of cartilage matrix protein within the von Willebrand factor type A-
RT like module superfamily.";
RL J. Biol. Chem. 272:9268-9274(1997)
CC -!- FUNCTION: INVOLVED IN MATRIX ASSEMBLY (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- TISSUE SPECIFICITY: DETECTED IN A VARIETY OF ORGANS, INCLUDING
CC CALVARIA, UTERUS, HEART, AND BRAIN, AS WELL AS FIBROBLAST AND
CC OSTEOBLAST CELL LINES.
CC -!- SIMILARITY: Contains 10 EGF-like domains.
CC -!- SIMILARITY: Contains 2 VWFA domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U69262; AAC53163.1;
DR HSP2; P05093; IAO5.
DR GDI; MG1109613; Matn2.
DR GO; GO:0005578; C:extracellular matrix; IDA.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR006209; EGF-Like.
DR InterPro; IPR006210; IEGF.
DR InterPro; IPR002035; VWFA.
DR Pfam; PF00008; EGF; 10.
DR PRINTS; PF00092; Vwa; 2.
DR PRINTS; PR00453; VWFADOMAIN.
DR SMART; SMC0181; EGF; 10.
DR SMART; SMC0127; VWA; 2.
DR PROSITE; PS00010; ASX_HYDROXYL; 7.
DR PROSITE; PS01186; EGF; 2; 9.
DR PROSITE; PS02034; VWFA; 2.
KW EGF-like domain; Signal; Glycoprotein; Repeat; Coiled coil.
FT SIGNAL 1 23 POTENTIAL.
FT CHAIN 24 956 MATRILIN-2.
FT DOMAIN 57 232 VWFA 1.
FT DOMAIN 238 278 EGF-LIKE 1.
FT DOMAIN 279 319 EGF-LIKE 2.
FT DOMAIN 320 360 EGF-LIKE 3.
FT DOMAIN 361 401 EGF-LIKE 4.
FT DOMAIN 402 442 EGF-LIKE 5.
FT DOMAIN 443 483 EGF-LIKE 6.
FT DOMAIN 484 524 EGF-LIKE 7.
FT DOMAIN 525 565 EGF-LIKE 8.
FT DOMAIN 566 606 EGF-LIKE 9.
FT DOMAIN 607 647 EGF-LIKE 10.
FT DOMAIN 655 830 VWFA 2.
FT DOMAIN 917 955 COILED COIL (POTENTIAL).
FT DISULFID 242 253 BY SIMILARITY.
FT DISULFID 249 262 BY SIMILARITY.
FT DISULFID 264 277 BY SIMILARITY.
FT DISULFID 283 294 BY SIMILARITY.
FT DISULFID 290 303 BY SIMILARITY.
FT DISULFID 305 318 BY SIMILARITY.
FT DISULFID 324 335 BY SIMILARITY.
FT DISULFID 331 344 BY SIMILARITY.
FT DISULFID 346 359 BY SIMILARITY.
FT DISULFID 365 376 BY SIMILARITY.
FT DISULFID 372 385 BY SIMILARITY.
FT DISULFID 387 400 BY SIMILARITY.
FT DISULFID 406 417 BY SIMILARITY.
FT DISULFID 423 426 BY SIMILARITY.
FT DISULFID 428 441 BY SIMILARITY.
FT DISULFID 447 458 BY SIMILARITY.
FT DISULFID 454 467 BY SIMILARITY.

```

```

FT DISULFID 469 482 BY SIMILARITY.
FT DISULFID 488 499 BY SIMILARITY.
FT DISULFID 495 508 BY SIMILARITY.
FT DISULFID 510 523 BY SIMILARITY.
FT DISULFID 529 540 BY SIMILARITY.
FT DISULFID 536 549 BY SIMILARITY.
FT DISULFID 551 564 BY SIMILARITY.
FT DISULFID 570 581 BY SIMILARITY.
FT DISULFID 577 590 BY SIMILARITY.
FT DISULFID 592 605 BY SIMILARITY.
FT DISULFID 611 622 BY SIMILARITY.
FT DISULFID 618 631 BY SIMILARITY.
FT DISULFID 633 646 BY SIMILARITY.
FT CARBOHYD 221 221 N-LINKED (GICNAC...) (POTENTIAL).
FT CARBOHYD 890 890 N-LINKED (GICNAC...) (POTENTIAL).
SQ SEQUENCE 956 AA; 106779 MW; 384C22770B6E6EE CRC64;

Query Match 12.3%; Score 692; DB 1; Length 956;
Best Local Similarity 29.8%; Pred. No. 1.3e-38;
Matches 159; Conservative 72; Mismatches 224; Indels 78; Gaps 14;

QY 62 QNTPTSYKCSKPGY--QGEGRQCEIDECNELNGCVHDCNIPNYRCTCFDGMFL 119
DB 253 CLNTPGSYICKQKQGYILSDQKTRIDLCATE-DHGCEQLCVNMLGFSVCQYGYTL 311
QY 120 AHGHNCLDVBCELENGGCOHTCVNMGSEYCCCKEGFFLSDNOHTCIHSEGLSCMN 179
DB 312 AEDGKRECTAMDCASENHGCEHCENASSYLRCHEGFALNSDKTKSKID----YCAS 367
QY 180 KDHCSSHCICKEAPRGSVACECPGPELAKNQRDC--ILTCNRHNGGCGHSCDDTDADPEC 237
DB 368 SNHGCOHECVNA-QTSALCRCLGFMFLNPDRTCRINVCALNKPGECEHCVENTEEGHVC 426
QY 238 SCHPOYKMTDGRSCLG-----REDTVLEVTESNTTSVVG-----DKRVKRLME 284
DB 427 RCRQGNLNPNGKTCRSVDHCAQDHGCEQLCLNTEESFVQCSEGF-INDELKTCRAQ 486
QY 285 TCANVNGGCDRTCKDTSTGVHSCVPGFTLQDGTCKDIDBCQFRNGGCGHFCFKNVGS 344
DB 487 YCLLSNHGCEYCVNTDSKSFACQCEGHVLRSDGKTKAKLSGALDGHGCEHSCVSSDS 546
QY 345 FDCGCKKGLKLTDEKSCQVDEC-SJDRTCDSHCINHPGTFACACNRYTL-YGFTHCG 402
DB 547 FVQCQFEGVILRDGKTCRRKQVQCVNHNGBHLGVNSGESVYCKLGLGFLAEDGKRCR 606
QY 403 C-NEGSINNGGCGQVQVNTVGSVECOCHPGYKLNKXQCV-----444
DB 607 RKNVCKSTQHGCEHMCVNGNNGNLYLCRSEGFVLAEDGKRCRGTGSPIDLVFVIDGSKS 666
QY 445 -----YKGLUPT-SVSPRYSLHCGKSGGDDGCFLRCHSGIHLSSEDTTIRTSVT 492
DB 667 GEENFETVGHFTGIIDS-LAVSPKA-----ARVQLQYSTQVTRTEFT 708
QY 493 FK-----LNEGKSLKVAEFPPEGLPALPEKUSSVKESFRYVNLTCSSGKQVP 542
DB 709 LRGFSASAKYKXAVTHMKYMGKSGMTGLAKH-MFERSFTQEGARPPESTQVF 760

```

RESULT 4

```

ID FBN1_BOVIN STANDARD; PRT: 2871 AA.
AC F98133;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Fibrillin 1 precursor (MF340).
GN FBN1.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID:9913;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC TISSUE=SKIN;
RX MEDLINE=95137597; PubMed=7835900;
RA Tilstra D.J., Potter K.A., Byers P.H.;
RT "Sequence of the coding region of the bovine fibrillin cDNA and
RL Genomics 23:480-485(1994).";
RN [2];
RP PARTIAL SEQUENCE.
RX MEDLINE=96132851; PubMed=8557636;
RA Gibson M.A., Hatzinikolas G., Kumaratilake J.S., Sandberg L.B.,
RA Nicholl J.K., Sutherland G.R., Cleary E.G.;
RT "Further characterization of proteins associated with elastic fiber
RL J. Biol. Chem. 271:1056-1063(1996).";
CC -/- FUNCTION: STRUCTURAL COMPONENT OF CONNECTIVE TISSUE MICROFIBRILS
CC THAT BINDS CALCIUM. FIBRILLIN-1-CONTAINING MICROFIBRILS PROVIDE
CC LONG-TERM FORCE BEARING STRUCTURAL SUPPORT.
CC -/- PTM: FORMS INTERMOLECULAR DISULFIDE BONDS EITHER WITH OTHER
CC FIBRILLIN-1 MOLECULES OR WITH OTHER COMPONENTS OF THE
CC MICROFIBRILS.
CC -/- SIMILARITY: Contains 47 EGF-like domains.
CC -/- SIMILARITY: Contains 7 TGF-beta binding protein (TGFBP) domains.
CC This SW-SS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EXBL: L28748; AAA74122...
CC PIR: A55567; A55567.
CC HSP: P35555; IAPU.
CC InterPro: IPR00152; Asx hydroxyl.
CC InterPro: IPR001881; EGF_CA.
CC InterPro: IPR006269; EGF_Like.
CC InterPro: IPR002212; Fibril-associ.
CC Pfam: PF00008; EGF; 46.
CC Pfam: PF00683; TB; 5.
CC SMART: SM00179; EGF_CA; 42.
CC PROSITE: PS00010; ASX HYDROXYL; 43.
CC PROSITE: PS00022; EGF_L; 2.
CC PROSITE: PS01186; EGF_L; 38.
CC PROSITE: PS01187; EGF_CA; 43.
CC Extracellular matrix; Calcium-binding; Glycoprotein; EGF-like domain;
KW Repeat; Signal; Multigene family.
FT SIGNAL 1 27 POTENTIAL.
FT CHAIN 26 2871 FIBRILLIN 1.
FT DOMAIN 81 112 EGF-LIKE 1.
FT DOMAIN 115 146 EGF-LIKE 2.
FT DOMAIN 147 178 EGF-LIKE 3.
FT DOMAIN 246 287 EGF-LIKE 4.
FT DOMAIN 288 329 EGF-LIKE 5.
FT REPEAT 330 390 TGFBP 1.
FT DOMAIN 392 446 PRO-RICH.
FT DOMAIN 449 489 EGF-LIKE 6.
FT DOMAIN 490 529 EGF-LIKE 7.
FT DOMAIN 530 571 EGF-LIKE 8.
FT DOMAIN 572 612 EGF-LIKE 9.
FT DOMAIN 613 653 EGF-LIKE 10.
FT REPEAT 654 722 TGFBP 2.
FT DOMAIN 723 764 EGF-LIKE 11.
FT REPEAT 765 806 EGF-LIKE 12.
FT DOMAIN 807 846 EGF-LIKE 13.
FT DOMAIN 847 952 EGF-LIKE 14.
FT REPEAT 952 1027 TGFBP 3.
FT DOMAIN 1028 1069 EGF-LIKE 15.
FT DOMAIN 1070 1112 EGF-LIKE 16.
FT DOMAIN 1113 1154 EGF-LIKE 17.
FT DOMAIN 1155 1196 EGF-LIKE 18.
FT DOMAIN 1197 1237 EGF-LIKE 19.

```

```
FT DOMAIN 1238 1279 EGF-LIKE 20, CALCIUM-BINDING.
FT DOMAIN 1280 1321 EGF-LIKE 21, CALCIUM-BINDING.
FT DOMAIN 1322 1362 EGF-LIKE 22, CALCIUM-BINDING.
FT DOMAIN 1363 1403 EGF-LIKE 23, CALCIUM-BINDING.
FT DOMAIN 1404 1445 EGF-LIKE 24, CALCIUM-BINDING.
FT DOMAIN 1446 1486 EGF-LIKE 25, CALCIUM-BINDING.
FT DOMAIN 1487 1527 EGF-LIKE 26, CALCIUM-BINDING.
FT REPEAT 1528 1605 TGFBP 4.
FT DOMAIN 1606 1647 EGF-LIKE 27, CALCIUM-BINDING.
FT DOMAIN 1648 1688 EGF-LIKE 28, CALCIUM-BINDING.
FT REPEAT 1689 1765 TGFBP 5.
FT DOMAIN 1766 1807 EGF-LIKE 29, CALCIUM-BINDING.
FT DOMAIN 1808 1848 EGF-LIKE 30, CALCIUM-BINDING.
FT DOMAIN 1849 1890 EGF-LIKE 31, CALCIUM-BINDING.
FT DOMAIN 1891 1929 EGF-LIKE 32, CALCIUM-BINDING.
FT DOMAIN 1930 1972 EGF-LIKE 33, CALCIUM-BINDING.
FT DOMAIN 1973 2012 EGF-LIKE 34, CALCIUM-BINDING.
FT DOMAIN 2013 2054 EGF-LIKE 35, CALCIUM-BINDING.
FT REPEAT 2055 2126 TGFBP 6.
FT DOMAIN 2127 2165 EGF-LIKE 36, CALCIUM-BINDING.
FT DOMAIN 2166 2205 EGF-LIKE 37, CALCIUM-BINDING.
FT DOMAIN 2206 2246 EGF-LIKE 38, CALCIUM-BINDING.
FT DOMAIN 2247 2290 EGF-LIKE 39, CALCIUM-BINDING.
FT DOMAIN 2291 2332 EGF-LIKE 40, CALCIUM-BINDING.
FT REPEAT 2333 2401 TGFBP 7.
FT DOMAIN 2402 2443 EGF-LIKE 41, CALCIUM-BINDING.
FT DOMAIN 2444 2484 EGF-LIKE 42, CALCIUM-BINDING.
FT DOMAIN 2485 2523 EGF-LIKE 43, CALCIUM-BINDING.
FT DOMAIN 2524 2566 EGF-LIKE 44, CALCIUM-BINDING.
FT DOMAIN 2567 2606 EGF-LIKE 45, CALCIUM-BINDING.
FT DOMAIN 2607 2647 EGF-LIKE 46, CALCIUM-BINDING.
FT DOMAIN 2648 2687 EGF-LIKE 47, CALCIUM-BINDING.
FT DISULFID 85 94
FT DISULFID 89 100
FT DISULFID 102 111
FT DISULFID 119 129
FT DISULFID 123 134
FT DISULFID 136 145
FT DISULFID 150 160
FT DISULFID 154 166
FT DISULFID 168 177
FT DISULFID 250 262
FT DISULFID 257 271
FT DISULFID 273 286
FT DISULFID 292 304
FT DISULFID 299 313
FT DISULFID 315 328
FT DISULFID 453 465
FT DISULFID 460 474
FT DISULFID 476 488
FT DISULFID 494 504
FT DISULFID 499 513
FT DISULFID 515 528
FT DISULFID 534 546
FT DISULFID 541 555
FT DISULFID 557 570
FT DISULFID 576 587
FT DISULFID 582 596
FT DISULFID 598 611
FT DISULFID 617 628
FT DISULFID 623 637
FT DISULFID 639 652
FT DISULFID 727 739
FT DISULFID 734 748
FT DISULFID 750 763
FT DISULFID 769 781
FT DISULFID 776 790
FT DISULFID 792 805
FT DISULFID 811 821
FT DISULFID 816 830
FT DISULFID 832 845
FT DISULFID 914 926
FT DISULFID 921 935
```

```
FT DISULFID 937 950 BY SIMILARITY.
FT DISULFID 1032 1044 BY SIMILARITY.
FT DISULFID 1039 1053 BY SIMILARITY.
FT DISULFID 1055 1068 BY SIMILARITY.
FT DISULFID 1074 1086 BY SIMILARITY.
FT DISULFID 1081 1095 BY SIMILARITY.
FT DISULFID 1097 1111 BY SIMILARITY.
FT DISULFID 1117 1129 BY SIMILARITY.
FT DISULFID 1124 1138 BY SIMILARITY.
FT DISULFID 1140 1153 BY SIMILARITY.
FT DISULFID 1159 1171 BY SIMILARITY.
FT DISULFID 1166 1180 BY SIMILARITY.
FT DISULFID 1182 1195 BY SIMILARITY.
FT DISULFID 1201 1212 BY SIMILARITY.
FT DISULFID 1208 1221 BY SIMILARITY.
FT DISULFID 1223 1236 BY SIMILARITY.
FT DISULFID 1242 1254 BY SIMILARITY.
FT DISULFID 1249 1263 BY SIMILARITY.
FT DISULFID 1265 1278 BY SIMILARITY.
FT DISULFID 1284 1296 BY SIMILARITY.
FT DISULFID 1291 1305 BY SIMILARITY.
FT DISULFID 1307 1320 BY SIMILARITY.
FT DISULFID 1326 1339 BY SIMILARITY.
FT DISULFID 1333 1348 BY SIMILARITY.
FT DISULFID 1350 1361 BY SIMILARITY.
FT DISULFID 1367 1380 BY SIMILARITY.
FT DISULFID 1374 1389 BY SIMILARITY.
FT DISULFID 1391 1402 BY SIMILARITY.
FT DISULFID 1408 1420 BY SIMILARITY.
FT DISULFID 1415 1429 BY SIMILARITY.
FT DISULFID 1431 1444 BY SIMILARITY.
FT DISULFID 1450 1461 BY SIMILARITY.
FT DISULFID 1456 1470 BY SIMILARITY.
FT DISULFID 1472 1485 BY SIMILARITY.
FT DISULFID 1491 1502 BY SIMILARITY.
FT DISULFID 1497 1511 BY SIMILARITY.
FT DISULFID 1513 1526 BY SIMILARITY.
FT DISULFID 1610 1622 BY SIMILARITY.
FT DISULFID 1617 1631 BY SIMILARITY.
FT DISULFID 1633 1646 BY SIMILARITY.
FT DISULFID 1652 1663 BY SIMILARITY.
```

Query Match 12.3%; Score 691; DB 1; Length 2871;

Best Local Similarity 22.7%; Pred. No. 5e-38;

Matches 271; Conservative 107; Mismatches 375; Indels 440; Gaps 49;

```
QY 45 DVDECAOGLDDCHADALCONPTSYKCSCKPGYOGEGROCEDIDEGKNEIANGGCVHDCLN 104
DB 1363 DLDECSNGTHKCSQHADCKNTMGSYKLCCKEGYTGDFCTDLDECSNENLNLGCGQCCLN 1422
QY 105 IPGNRYCTCFDGFMLAHDHGHNCLDVDEC-----LENNNG 137
DB 1423 APGGYRCCECMGFVPSADGKACEDIDECSLPNICVFETGCHNLPGLFRCCEIGYELDRSG 1482
QY 138 G-----COHCTCVNMGVSECCKEGFELSDNCHTCH-----S 171
DB 1493 GNCTDVNECLDPTTCISGNVCNVTGPGSTYCDPPDFELNTRVCGVDTNRSGNVCYLDIRPG 1542
QY 172 EEG-LSCMN-----KHGCSHI----- 187
DB 1543 DNGDTACSNEIGVGVSKASCCSLGKAWGTPCELCPVNTSEYKILCPGEGEPNPITV 1602
QY 188 -----CKEAPR-----GSVACECRPGFELAKNQKQDC--ILTCNH-----GNCG 223
DB 1603 ILEIDECQELPGLCGGKGCINFGSPQRCRCPTGYVLNEDTRVCCDVNCECTPGICGPG- 1661
QY 224 CQHSCDCTAGPESCHPQYKMHCDGSGSLERSDTIVLEVTESNTTSVVDGCKEYKRLIM 283
DB 1662 ---TCYNTGVNYTCIGPPDYMQVNGGNNCMDMRSJCYRNYVADNCTCGELLFNTKMM 1718
QY 284 ETCAVNNG-GCDRTCK-----DTSTGV----- 304
DB 1719 CCSSYNIGRANWKPCEOCFIPSTDEATCGSORPGFVIDITGLPVDIDECRETPGVCE 1778
```

```
QY 305 -----HCSPVGTFLQDKTKDIDBCQ----- 328
Db 1779 NGVCINMWSFRCECPVGFYNDKLLVCEDIDECQNGPVQORNAECINTAGSYRCCCKPG 1838
QY 329 ---TRNGGDH-----FCNKIVGDFCGCKKFKLLTDEKSCQDVDECSLD 371
Db 1839 YRFTSTGQCDNRNECOEIPNCSHGOCIDTVGSFYCLHTGKTNADQTMCLDINECED 1898
QY 372 RTCDHSCINHPGTFACNRGTYLXGTHCGDTNECSINNGG-CQ-QVNVTVGSYECQC 429
Db 1899 ACNGTCRNTIGSFNCRNHFILSHNDICIDVDECATGNLCRNGQCINTVGSFQCQC 1958
QY 430 HPGYKLMHKKQDVEYKGLL---PTSVPRVSLHCKSGGGDCFLRCHSHILSDVTT 487
Db 1959 NEGVEVAPDGTCDVINECLLDRKCAPGT---CQNLDSYRCI---CPFGVSLQNDKCE- 2012
QY 488 RTSVTFKLME-----GKCSLKNAE-----LPPEGLR-PALPEKHSSVKESFRYN 531
Db 2013 -----DIDECVEEPEICALGTCS---NTEGSPKCLCPDGFSLSTGRRCQDLKSYCYAK 2064
QY 532 LTCSSKQVPGAGRSTRKEMFIVPELETNQKEVTASCDSL-----CIVKRTSKR 584
Db 2065 F-----EGKCSSPK-----SRHKGQCECCALKGEGWDGDCCLCPTSF- 2103
QY 585 LRKAIRTLKAVHREOFHLQL-----SQMLLVAKKPPRTSERQAES----- 626
Db 2104 -----DEAFQICPYQSGIIVGDDSAVDECKEFDVKHGKCINTDSY 2149
QY 627 ---CGVQGHAEQVQ---SCRAGTYDGARERCILCPNGTQNEBQWTEPCPRPONS 68C
Db 2150 RCECPGYLQNECVDTDECSVGN-----PCNGTCKNVIGSFEC-TCBEGSEP 2198
QY 681 GALKTEPANNMECCG---LCQPGESADGAPCOLCALGTQPEAGRTSC---FPCGG 734
Db 2199 GPMTCF---DINECAONPLLCAPRCNTVGSVECK-CPAGYVLR-DRMCKDECESEG 2254
QY 735 LATKHQATSFQDCEFRVQ-----CSPGHFYNTTHRCIRCVPVGTQYQPFQKNCVSCPG 789
Db 2255 ---KHCEAKQWCKEKNL-GTYL-CICPGYQRRPDGEGCV-----DENECQCKPS 2320
QY 790 NTTDFEGSTNTQCKNRCCGELGDT-----GYIESPNYGNYPANTECTWTIN?PPK 844
Db 2301 -----ICENGRCLNTRGTYCECDGETASEN----- 2327
QY 845 RRLIVVPEILFPIEDCGD-----YLVVRKTSNNSVNTYETCTQYER--- 886
Db 2328 -----QDECLDNREGYCFTEVLQNMCOIGSSNRNPVTKSECCCGRGKWP 2373
QY 889 -----PIAFTRSKLWLOFKSNEGNSARGFQVPVYVYEDYQELLEDIVRDG 936
Db 2374 HCEICPFQGTVPFKLCL-----PHGRGFTNGAD:DE--CKVIHDVCRNG 2416

RESULT 5
FBN1_HUMAN
ID FBN1_HUMAN STANDARD; PRT; 2871 AA.
AC P35555;
DT 01-JUN-1994 (Rel. 29, Created)
DI 01-JUN-1994 (Rel. 29, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Fibrillin 1 precursor.
GN FBN1 OR FBN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=93372860; PubMed=8364578;
RA Pereira L.V., D'Alessio M., Ramirez F., Lynch J.R., Sykes B.,
RA Pangilinan T., Bonadio J.;
```

```
RT "Genomic organization of the sequence coding for fibrillin, the
RL defective gene product in Marfan syndrome.";
RN Hum. Mol. Genet. 2:961-968(1993).
RP [2]
RP SEQUENCE OF 1-932 FROM N.A.
RC T-SSUE=Fibroblast, and Placenta;
RX MEDLINE=9401C947; PubMed=7691719;
RA Corson G.M., Chabberg S.C., Dietz H.C., Charbonneau N.J., Sakai L.Y.;
RT "Fibrillin binds calcium and is coded by cDNAs that reveal a
RT multidomain structure and alternatively spliced exons at the 5'
end.";
RN Genomics 17:476-484(1993).
RP [3]
RP SEQUENCE OF 899-2871 FROM N.A.
RX MEDLINE=91304567; PubMed=8522206;
RA Masien C.L., Corson G.M., Maddox B.K., Glanville R.W., Sakai L.Y.;
RC "Partial sequence of a candidate gene for the Marfan syndrome.";
RN Nature 352:334-337(1991).
RP [4]
RP SEQUENCE OF 813-1313 FROM N.A.
RX MEDLINE=91304567; PubMed=8522206;
RA Lee S., Godfrey M., Vitale E., Horii H., Mattei M.-G., Sarfarazi M.,
RA Tsipouras P., Ramirez F., Hollister C.W.;
RT "Linkage of Marfan syndrome and a phenotypically related disorder to
RT two different fibrillin genes.";
RN Nature 352:330-334(1991).
RP [5]
RP CHARACTERIZATION.
RX MEDLINE=91317849; PubMed=860873;
RA Sakai L.Y., Keene D.R., Glanville R.W., Bachinger H.P.;
RT "Purification and partial characterization of fibrillin, a cysteine-
RT rich structural component of connective tissue microfibrils.";
RN J. Biol. Chem. 266:14763-14770(1991).
RP [6]
RP STRUCTURE BY NMR OF 2054-2225.
RX MEDLINE=96031893; PubMed=8562480;
RA Yuan X., Downing A.K., Knott V., Handford P.A.;
RT "Solution structure of the transforming growth factor beta-binding
RT protein-like module, a domain associated with matrix fibrils.";
RN EMBO J. 16:6659-6666(1997).
RP [7]
RP STRUCTURE BY NMR OF 2124-2285.
RX MEDLINE=96144829; PubMed=8568869;
RA Knott V., Downing A.K., Cady C.M., Handford P.A.;
RT "Calcium binding properties of an epidermal growth factor-like domain
RT pair from human fibrillin-1.";
RN J. Mol. Biol. 255:22-27(1996).
RP [8]
RP STRUCTURE BY NMR OF 2124-2205.
RX MEDLINE=96222301; PubMed=8653784;
RA Downing A.K., Knott V., Werner J.N., Cady C.M., Campbell I.D.,
RA Handford P.A.;
RT "Solution structure of a pair of calcium-binding epidermal growth
RT factor-like domains: implications for the Marfan syndrome and other
RT genetic disorders.";
RN Cell 85:597-605(1996).
RP [9]
RP REVIEW ON MFS VARIANTS.
RX MEDLINE=96174615; PubMed=8594563;
RA Colod G., Beroud C., Soussi T., Gunten C., Boileau C.;
RT "Software and database for the analysis of mutations in the human
RT FBN1 gene.";
RN Nucleic Acids Res. 24:137-141(1996).
RP [10]
RP REVIEW ON MFS VARIANTS.
RX MEDLINE=97169383; PubMed=9016526;
RA Colod-Beroud G., Beroud C., Ades L., Slack C., Boxer M., Brock D.C.,
RA Godfrey X., Hayward C., Karltun L., Milewicz D., Peltonen L.,
RA Richards R.I., Wang M., Gunten C., Boileau C.;
RT "Marfan Database (second edition): software and database for the
RT analysis of mutations in the human FBN1 gene.";
RN Nucleic Acids Res. 25:147-150(1997).
RP [11]
```

RP REVIEW ON VARIANTS.
 RX MEDLINE=98062175; PubMed=9401003;
 RA Hayward C., Brock D.J.H.;
 RT "Fibrillin-1 mutations in Marfan syndrome and other type-1
 RL fibrillinopathies";
 RL Hum. Mutat. 10:415-423 (1997).
 RN [12]
 RP VARIANT MFS PRO-1137.
 RX MEDLINE=91304569; PubMed=1852208;
 RA Dietz H.C., Cutting G.R., Pyeritz R.E., Maslen C.L., Sakai L.Y.,
 RA Corson G.M., Puffenberger E.G., Hamosh A., Nanthakumar E.J.,
 RA Curristin S.M., Stetten G., Meyers D.A., Francomano C.A.;
 RT "Marfan syndrome caused by a recurrent de novo missense mutation in
 RL the fibrillin gene";
 RL Nature 352:337-339 (1991).
 RN [13]
 RP VARIANTS MFS SER-1249; ARG-1663; SER-2221 AND SER-2307.
 RX MEDLINE=93250834; PubMed=1301946;
 RA Dietz H.C., Saraiva J.M., Pyeritz R.E., Cutting G.R., Francomano C.A.;
 RT "Clustering of fibrillin (FBN1) missense mutations in Marfan syndrome
 RL patients at cysteine residues in EGF-like domains";
 RL Hum. Mutat. 1:366-374 (1992).
 RN [14]
 RP VARIANT MFS SER-2307.
 RX MEDLINE=92235290; PubMed=1569206;
 RA Dietz H.C., Pyeritz R.E., Puffenberger E.G., Kendzior R.J. Jr.,
 RA Corson G.M., Maslen C.L., Sakai L.Y., Francomano C.A., Cutting G.R.;
 RT "Marfan phenotype variability in a family segregating a missense
 RL mutation in the epidermal growth factor-like motif of the fibrillin
 RT gene";
 RL J. Clin. Invest. 89:1674-1680 (1992).
 RN [15]
 RP VARIANTS MFS ILE-548 AND ALA-723.
 RX MEDLINE=94010946; PubMed=8406497;
 RA Dietz H.C., McIntosh I., Sakai L.Y., Corson G.M., Chalberg S.C.,
 RA Pyeritz R.E., Francomano C.A.;
 RT "Four novel FBN1 mutations: significance for mutant transcript level
 RL and EGF-like domain calcium binding in the pathogenesis of Marfan
 RT syndrome";
 RL Genomics 17:468-475 (1993).
 RN [16]
 RP VARIANTS MFS SER-2144.
 RX MEDLINE=93278402; PubMed=8504310;
 RA Hewett D.R., Lynch J.R., Smith R., Sykes B.C.;
 RT "A novel fibrillin mutation in the Marfan syndrome which could
 RL disrupt calcium binding of the epidermal growth factor-like module";
 RL Hum. Mol. Genet. 2:475-477 (1993).
 RN [17]
 RP VARIANTS MFS ARG-862; TYR-1117; PRO-1137 AND PHE-1589, AND VARIANT
 RL ALA-1148.
 RX MEDLINE=94108431; PubMed=8281141;
 RA Tynan K., Comeau K., Pearson M., Wilgenbus P., Levitt D., Gasner C.,
 RA Berg M.A., Miller D.C., Francke U.;
 RT "Mutation screening of complete fibrillin-1 coding sequence: report
 RL of five new mutations, including two in 8-cysteine domains";
 RL Hum. Mol. Genet. 2:1813-1821 (1993).
 RN [18]
 RP VARIANTS MFS GLY-217 AND ARG-2627.
 RX MEDLINE=95067970; PubMed=7977366;
 RA Karttunen L., Raghunath M., Lonnqvist L.;
 RT "A compound-heterozygous Marfan patient: two defective fibrillin
 RL alleles result in a lethal phenotype";
 RL Am. J. Hum. Genet. 55:1083-1091 (1994).
 RN [19]
 RP VARIANT EL LYS-2447.
 RX MEDLINE=94245249; PubMed=8188302;
 RA Lonnqvist L., Child A., Kainulainen K., Davidson R., Puhakka L.,
 RA Peltonen L.;
 RT "A novel mutation of the fibrillin gene causing ectopia lentis";
 RL Genomics 19:573-576 (1994).
 RN [20]
 RP VARIANT MFS CYS-627.
 RX MEDLINE=94272487; PubMed=8004112;

RA Hayward C., Rae A.L., Porteous M.E.M., Logie L.J., Brock L.J.;
 RT "Two novel mutations and a neutral polymorphism in EGF-like domains
 RL of the fibrillin gene (FBN1): SSCP screening of exons 15-21 in Marfan
 RL syndrome patients";
 RL Hum. Mol. Genet. 3:373-375 (1994).
 RN [21]
 RP VARIANT MFS CYS-122.
 RX MEDLINE=94314977; PubMed=8040326;
 RA Stahn-Hallengren C., Ukkonen T., Kainulainen K., Kristofersson U.,
 RA Saxne T., Tornqvist K., Peltonen L.;
 RT "An extra cysteine in one of the non-calcium-binding epidermal growth
 RL factor-like motifs of the FBN1 polypeptide is connected to a novel
 RL variant of Marfan syndrome";
 RL J. Clin. Invest. 94:709-713 (1994).
 RN [22]
 RP VARIANT MFS TYR-1223.
 RX MEDLINE=94351682; PubMed=8071963;
 RA Hewett D.R., Lynch J.R., Child A., Sykes B.C.;
 RT "A new missense mutation of fibrillin in a patient with Marfan
 RL syndrome";
 RL J. Med. Genet. 31:339-339 (1994).
 RN [23]
 RP VARIANT MFS HIS-117C.
 RX MEDLINE=95174777; PubMed=7870075;
 RA Hayward C., Porteous M.E.M., Brock D.J.H.;
 RT "A novel mutation in the fibrillin gene (FBN1) in familial
 RL atachnodyly";
 RL Mol. Cell. Probes 8:325-327 (1994).
 RN [24]
 RP VARIANTS MFS GLY-217; ASN-1623; ARG-1074; TYR-1242; ARG-1513;
 RL GLY-2127; TRP-2151; LYS-2447 AND ARG-2511.
 RX MEDLINE=94184368; PubMed=8136837;
 RA Kainulainen K., Karttunen L., Puhakka L., Sakai L., Peltonen L.;
 RT "Mutations in the fibrillin gene responsible for dominant ectopia
 RL lentis and neonatal Marfan syndrome";
 RL Nat. Genet. 6:64-69 (1994).
 RN [25]
 Query Match 12.3%; Score 691; DB 1; Length 2871;
 Best Local Similarity 22.5%; Pred. No. 5e-38;
 Matches 269; Conservative 105; Mismatches 378; Indels 442; Gaps 45;
 Qy 45 DVDECAAGLDDCHADALCQNTPTSYKSCKPKVQGGROCEDIDECGNELNGSCVHDCLN 134
 Db 1363 DLDECSNGTHVCSQHADCKNTGVSRLCKEGYTGSGFTCTDLDECSNENLNLGNGQCLN 1422
 Qy 105 LPGNYRCTCFDGFMLADHGNCLDVECLNNGCGCHTCVNVVMSVECCCKEGFFLSDNQ 164
 Db 1423 AFGYRCECDMGVPSADGKACEDIDECSLPNLCVFGTCHNLPGLPRCECEIGYELDRSG 1482
 Qy 165 HCTIHRSEEGLSQMKXDHGCSHCKSAPRGSVACRPGFELAKNQDCLLT----- 216
 Db 1483 GNCTDVNE----CLDPTCTISGNCVNTPT-GSYICDCPPDFELNPTRVGCVDRSGNCLYD 1537
 Qy 217 ----- 216
 Db 1538 IRPRGNGDTACSNETIGVGVSKASCCSCGKAWGTPCEMCPAVNTSEYKILCPGGSGFRP 1597
 Qy 217 -----CNHNGNGGC-----HSCDDTAD----- 233
 Db 1598 NPITVILEIDIEQELPGLCQGGKCNINTFGSPQCRCTGYVLNEDTRVCDVNECETPGT 1657
 Qy 234 -GP-----ECSCHPQYKQHTDGRSCLEREDTVLEVTSTNTTSSVVDGKXKVRLL 282
 Db 1658 CGPQTCVNTGVNVTCTCPPYQVQVNGNNGNMDVRSLSLCYRNVYADNQTCDGELLENTKK 1717
 Qy 293 METCAVNG--GCDRIK-----DTSTGV----- 304
 Db 1718 MCCCSYNIGRAMNKPTEOCPITPSTDEFATILCGSORPGFVIDYTGLPVIDIDECRE:PGVC 1777
 Qy 305 -----HCSGPVGETLQDKCTCKDIDECQ----- 328
 Db 1778 ENGVCINNVGSGFRCEPVGFFYNDKJLVCEIDIECCNGPVCGNACINTAGSYRCDCKP 1837


```

QY 329 -----TRNGGCDH-----FCNIVGSPDCCKGKFKLLTDEKSCQVDDECSL 370
DB 1838 GYRFTSTQCNDNRNCOBIPNICSHGQCIDTWSFYCLCHTGFKNDDQTNCLDINCEK 1837
QY 371 DRTCHSCINHPGTAFACNRYTLYGTHCGDTNECSINNG-CCQ-VCVNTVGSVECCQ 429
DB 1898 DACNGTCTRTVGTGNCNCRNGHFLSHNNDICIDVDEASGNVLCRNGCCINTVGSFCCQ 1957
QY 429 CHPGYKLHNNKDCVVEVKGLL--PTSVPRVSLHCGKSGGGGDFLCHSGIHLSSDVT 486
DB 1958 CNEGYEAPDGRCTVDINECLPRKCAPGT--CQNLDSGYRCI--CPCGYSLONEKCE 2012
QY 487 TRTSVTFKLINE-----GKSLKNAE-----LFPEG-R-PALPKHSVVESEFV 530
DB 2013 -----DIDECVEPEICALGTCS--NTEGSFKLCPEGFSLSGSSRRCCQLRMSVCYA 2063
QY 531 NLTSSGQKVPAGRPSTPKEMFTVFELETNCKEVTASCDLS-----CIVKXTEK 583
DB 2064 KF-----EGKCSPK-----SRNHSKQECCECCALKGSGWGPCELCCTEP 2103
QY 584 RLRKAIRTLRKAVHREQHQL-----SGYNLDVAKKPPRTEROABE----- 626
DB 2104 -----DEAFRQICPYGSGIIVGPDSDAVDMJCEKPPVCKKHGQCINTDGS 2148
QY 627 -----CGVGQGHAEKNCV-----SCRAGTYVDGAHERCILCPNGTTFONEEGQMTCEPCPRGN 579
DB 2149 YKCEPFGVTLJAGNECVDTDECSVN-----PCONG:CKNV:GGFEC-TCEBQFE 2197
QY 680 SGALKTPAANNMSEGG-----LCQGEYSADGAPQQLCALGTFCPEAGRTSC---FPCGG 733
DB 2198 PGPMMTCE-----DINECAQNPLLCAFRCVNTYGSYECK-CPVGVVLE-DRMCKDEDECEE 2253
QY 734 G-ATKHQGA:SFQDCETRVQ-----CSPGHFVNTTHRCIRCPVGTGYOPEFGKNCVSCP 788
DB 2254 G---KHDCTEKQMECKNLJGTMYCIGPGYQRRPDGEGCV-----DENECQTKP 2299
QY 789 GNTTTFDGSNTITOCNRRCGELGDET-----GVIESPNYPGNYPANTCTWTINPP 843
DB 2300 G-----ICENGRCLNTRGSYTCECNDGFTASPN----- 2327
QY 844 KRRILIVWEIPLIEDDCGD-----YLVMRKTSNSNVTVYTCQTVR--- 888
DB 2328 -----QDECLDNREGYCFTEVLQNMCIQSSNRNPNVTKSECCDGRGWG 2372
QY 889 -----PIAFTSRKKLWTFKNSARGQVPPVYDEYQJELIEDIYRDG 936
DB 2373 PHCE:CPFOGTVAFKKLC-----PHORGFMINGADIDE--CKVIHDVCNKG 2416

RESULT 6
ID FBN1_MOUSE STANDARD; PRT: 2871 AA.
AC Q61554; Q60826;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Fibrillin 1 precursor.
GN FBN1 OR FBN-1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95130561; PubMed=7829516;
RA Yin W., Gernier J., Sanguinetti C., Smiley S., Pangilinan T.,
RA Pereira L., Ramirez F., Bonadio J.;
RT "Primary structure and developmental expression of Fbn-1, the mouse
RT fibrillin gene.";
RL J. Biol. Chem. 270:1798-1806(1995).
RN [2]
RP SEQUENCE FROM N.A.

```

```

RC STRAIN=CD-1; TISSUE=Kidney;
RA Cta K., Kumar A., Wada J., Liu Z., Kanwar Y.S.;
RL Submitted (APR-1995) to the EMBL/Genbank/DBSD databases.
CC -!- FUNCTION: STRUCTURAL COMPONENT OF CONNECTIVE TISSUE MICROFIBRILS
CC THAT BINDS CALCIUM. FIBRILLIN-1-CONTAINING MICROFIBRILS PROVIDE
CC LONG-TERM FORCE BEARING STRUCTURAL SUPPORT.
CC -!- PTM: FORMS INTERMOLECULAR DISULFIDE BONDS EITHER WITH OTHER
CC FIBRILLIN-1 MOLECULES OR WITH OTHER COMPONENTS OF THE
CC MICROFIBRILS (BY SIMILARITY).
CC -!- SIMILARITY: Contains 47 EGF-like domains.
CC -!- SIMILARITY: Contains 7 TGF-beta binding protein (TGFBP) domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to licenses@isb-sib.ch).
CC
CC -----
CC ENBL; L29454; AAA56840.1; --
CC ENBL; U22493; AAA6427.1; --
CC PIR; A55624; A55624.
CC HSP; P35555; IAPQ.
CC MGC; MG1:95489; Fbn1.
CC InterPro: IPR000152; Asx_Hydroxyl-
CC InterPro: IPR00188; EGF_Ca
CC InterPro: IPR006209; EGF_Like.
CC InterPro: IPR002212; Fibril-assoc.
CC Pfam; PF00038; EGF; 46.
CC Pfam; PF0683; TB; 9.
CC SMART; SM00179; EGF_CA; 42.
CC PROSITE; PS00310; ASX_HYDROXYL; 43.
CC PROSITE; PS00322; EGF_1; 2.
CC PROSITE; PS01186; EGF_2; 38.
CC PROSITE; PS01187; EGF_CA; 43.
CC Extracellular matrix; Calcium-binding; Glycoprotein; EGF-like domain;
KW Repeat; Signal; Multigene family.
FT SIGNAL 1 27
FT STRAIN 28 2871
FT FIBRILLIN 1.
FT DOMAIN 8; 112
FT EGF-LIKE 1.
FT DOMAIN 115 146
FT EGF-LIKE 2.
FT DOMAIN 147 178
FT EGF-LIKE 3.
FT DOMAIN 246 287
FT EGF-LIKE 4.
FT DOMAIN 288 329
FT EGF-LIKE 5.
FT DOMAIN 330 401
FT TGFBP 1.
FT DOMAIN 402 446
FT PRO-RICH.
FT DOMAIN 449 489
FT EGF-LIKE 6.
FT DOMAIN 490 529
FT EGF-LIKE 7.
FT DOMAIN 530 571
FT EGF-LIKE 8.
FT DOMAIN 572 612
FT EGF-LIKE 9.
FT DOMAIN 613 653
FT EGF-LIKE 10.
FT DOMAIN 656 721
FT TGFBP 2.
FT DOMAIN 723 764
FT EGF-LIKE 11.
FT DOMAIN 765 806
FT EGF-LIKE 12.
FT DOMAIN 807 846
FT EGF-LIKE 13.
FT DOMAIN 910 951
FT EGF-LIKE 14.
FT DOMAIN 952 1018
FT TGFBP 3.
FT DOMAIN 1028 1069
FT EGF-LIKE 15.
FT DOMAIN 1070 1112
FT EGF-LIKE 16.
FT DOMAIN 1113 1154
FT EGF-LIKE 17.
FT DOMAIN 1155 1196
FT EGF-LIKE 18.
FT DOMAIN 1197 1237
FT EGF-LIKE 19.
FT DOMAIN 1238 1279
FT EGF-LIKE 20.
FT DOMAIN 1280 1321
FT EGF-LIKE 21.
FT DOMAIN 1322 1362
FT EGF-LIKE 22.
FT DOMAIN 1363 1403
FT EGF-LIKE 23.
FT DOMAIN 1404 1445
FT EGF-LIKE 24.
FT DOMAIN 1446 1486
FT EGF-LIKE 25.
FT DOMAIN 1487 1527
FT EGF-LIKE 26.
FT DOMAIN 1528 1569
FT TGFBP 4.
FT DOMAIN 1606 1647
FT EGF-LIKE 27.
FT DOMAIN 1648 1688
FT EGF-LIKE 28.

```

```
FT DOMAIN 1689 1758 TGFBR 5. EGF-LIKE 29, CALCIUM-BINDING. BY SIMILARITY.
FT DOMAIN 1766 1807 EGF-LIKE 30, CALCIUM-BINDING. BY SIMILARITY.
FT DOMAIN 1808 1848 EGF-LIKE 31, CALCIUM-BINDING. BY SIMILARITY.
FT DOMAIN 1849 1890 EGF-LIKE 32, CALCIUM-BINDING. BY SIMILARITY.
FT DOMAIN 1891 1932 EGF-LIKE 33, CALCIUM-BINDING. BY SIMILARITY.
FT DOMAIN 1930 1972 EGF-LIKE 34, CALCIUM-BINDING. BY SIMILARITY.
FT DOMAIN 1973 2012 EGF-LIKE 35, CALCIUM-BINDING. BY SIMILARITY.
FT DOMAIN 2013 2054 TGFBR 6. EGF-LIKE 36, CALCIUM-BINDING. BY SIMILARITY.
FT DOMAIN 2055 2121 EGF-LIKE 37, CALCIUM-BINDING. BY SIMILARITY.
FT DOMAIN 2127 2165 EGF-LIKE 38, CALCIUM-BINDING. BY SIMILARITY.
FT DOMAIN 2166 2205 EGF-LIKE 39, CALCIUM-BINDING. BY SIMILARITY.
FT DOMAIN 2206 2246 EGF-LIKE 40, CALCIUM-BINDING. BY SIMILARITY.
FT DOMAIN 2247 2290 TGFBR 7. EGF-LIKE 41, CALCIUM-BINDING. BY SIMILARITY.
FT DOMAIN 2291 2332 EGF-LIKE 42, CALCIUM-BINDING. BY SIMILARITY.
FT DOMAIN 2333 2400 EGF-LIKE 43, CALCIUM-BINDING. BY SIMILARITY.
FT DOMAIN 2402 2443 EGF-LIKE 44, CALCIUM-BINDING. BY SIMILARITY.
FT DOMAIN 2444 2484 EGF-LIKE 45, CALCIUM-BINDING. BY SIMILARITY.
FT DOMAIN 2485 2523 EGF-LIKE 46, CALCIUM-BINDING. BY SIMILARITY.
FT DOMAIN 2524 2566 EGF-LIKE 47, CALCIUM-BINDING. BY SIMILARITY.
FT DOMAIN 2567 2606 BY SIMILARITY.
FT DOMAIN 2607 2647 BY SIMILARITY.
FT DOMAIN 2648 2687 BY SIMILARITY.
FT DISULFID 85 94 BY SIMILARITY.
FT DISULFID 89 100 BY SIMILARITY.
FT DISULFID 102 111 BY SIMILARITY.
FT DISULFID 119 129 BY SIMILARITY.
FT DISULFID 123 134 BY SIMILARITY.
FT DISULFID 136 145 BY SIMILARITY.
FT DISULFID 150 160 BY SIMILARITY.
FT DISULFID 154 166 BY SIMILARITY.
FT DISULFID 168 177 BY SIMILARITY.
FT DISULFID 250 262 BY SIMILARITY.
FT DISULFID 257 271 BY SIMILARITY.
FT DISULFID 273 286 BY SIMILARITY.
FT DISULFID 292 304 BY SIMILARITY.
FT DISULFID 299 313 BY SIMILARITY.
FT DISULFID 315 328 BY SIMILARITY.
FT DISULFID 453 465 BY SIMILARITY.
FT DISULFID 460 474 BY SIMILARITY.
FT DISULFID 476 488 BY SIMILARITY.
FT DISULFID 494 504 BY SIMILARITY.
FT DISULFID 499 513 BY SIMILARITY.
FT DISULFID 515 528 BY SIMILARITY.
FT DISULFID 534 546 BY SIMILARITY.
FT DISULFID 541 555 BY SIMILARITY.
FT DISULFID 557 570 BY SIMILARITY.
FT DISULFID 576 587 BY SIMILARITY.
FT DISULFID 582 596 BY SIMILARITY.
FT DISULFID 598 611 BY SIMILARITY.
FT DISULFID 617 628 BY SIMILARITY.
FT DISULFID 623 637 BY SIMILARITY.
FT DISULFID 639 652 BY SIMILARITY.
FT DISULFID 727 739 BY SIMILARITY.
FT DISULFID 734 748 BY SIMILARITY.
FT DISULFID 750 763 BY SIMILARITY.
FT DISULFID 769 781 BY SIMILARITY.
FT DISULFID 776 790 BY SIMILARITY.
FT DISULFID 792 805 BY SIMILARITY.
FT DISULFID 811 821 BY SIMILARITY.
FT DISULFID 816 830 BY SIMILARITY.
FT DISULFID 832 845 BY SIMILARITY.
FT DISULFID 914 926 BY SIMILARITY.
FT DISULFID 921 935 BY SIMILARITY.
FT DISULFID 937 950 BY SIMILARITY.
FT DISULFID 1032 1044 BY SIMILARITY.
FT DISULFID 1039 1053 BY SIMILARITY.
FT DISULFID 1055 1068 BY SIMILARITY.
FT DISULFID 1074 1086 BY SIMILARITY.
FT DISULFID 1081 1095 BY SIMILARITY.
FT DISULFID 1097 1111 BY SIMILARITY.
FT DISULFID 1117 1129 BY SIMILARITY.
FT DISULFID 1124 1138 BY SIMILARITY.
FT DISULFID 1140 1153 BY SIMILARITY.
```

Query Match 12.1% Score 690; DB 1; Length 2871;
Best Local Similarity 23.4% Pred No. 5.8e-38;
Matches 276; Conservativeness 104; Mismatches 384; Indels 418; Gaps 51;

QY 45 DVDECAQGLDDCHADALCONTPTSYKCSCKPGYGGEGQCEIDECGNELNGGCVHCLN 104
DB 1363 DLDECSNGTHMCSQHADCKNTMGSYRCLCKDGVGTGDTCTDLDECSNLCNGGCLN 1422
QY 105 IPGNRYCTCDGFMALHDGHNCILDVDEG-----LENGG 137
DC 1423 APGGYRCEDYGFVPSADGKACEDIDECSPNTCVFGTCHNLPGLPFCCEIGYELDRSG 1482
QY 138 G-----CGHTCVNMGSYECCCKEGFFLENGHTCH-----RS 171
DB 1483 GNCTDVNECLDPTTC:SGNCVNTPGSYTCDGSPDFELNPTRVGCVDRSNCVNIAPRG 1542
QY 172 EELSCVNRKOHG-----CSHI----- 187
DB 1543 ENGDTACSNEIGVGVSKASCCSLGKAWGTFCBLCPSPVNTSEYKILCPGGEGRFNPTV 1602
QY 188 -----CKEAPR-----GSVACEGRGPELAKNQRDC--ILTCNH-----QNGG 223
DB 1603 ILEDIDECQELPGLCGGKGINTFGQCRCPTGYJNEJTRVCDVNEGETPGICGPG- 1661
QY 224 CQHSCTDTADGPECSCHPQYKMHITDGRASCLEREDTLEVTESNTTSVVDGDKRVRKRLLM 283
DB 1662 ---TCYNTVGNVTC:CPDYQVQVNGNNDKMRSLCYRNYADNCTCQGLL-FNMTKKY 1718
QY 284 ETCAN-----NGGCRCTCKDTT----- 302
DB 1719 CCSSYNIIRAWNKPEEQ-CPIPSDEPATLCSGSRPGFVIDYTGLPVDIDECREIPGVC 1777
QY 303 --GV-----HCSQVGGFTLOLDGKTCXGIDECQ----- 328
DB 1778 ENGVC:NMVGSFRCCEPVGFYNDKLLVQSD:DECCQNPVCLNAECINPAGSYRCDCKP 1837
QY 329 ----TRNGGCDH-----FCKNIVGSFDCGCKKFKJLTDEKSCQVDECSJ 370
DB 1838 GYRJTSTGQCNDNRNECQEIPIKSHGOCIDTVSGFYCLCHTGFTNFDQNCUSINECEP 1897

QY 371 DRTCHSCINHPCTPACACNROVITLYGTHGDTNECSINNG-CQ-QVCVNTVGSVECC 428
 Db 1998 DAGNGTCTRTIGSFNCRNMGHFIILSHNNDCIDVDGATGNLGRNQCQVNTVGSFQCR 1957
 QY 429 CHPGYKLHNKDCVEVKG--LLPTSVSRVSLHCKSGGGGDFLRCGSHGLHSSDVTT 486
 Db 1958 CNEGEVADPGTCDVINECVLDPOKCAPGT---CONLDGVRCI--CPPGVSLNDKCE 2012
 QY 487 IRTSVTFKLINE-----GKSLKNAE-----LPPEGLR-PALPEKSHKSVKESPRV 530
 Db 2013 -----DIDECVEPEICALGTCG--NTEGSFKLCPEGFSWSSGRRRCQDLRXYCYA 2063
 QY 531 NUTCSGKQVPCAPRBPSTPKEMFIVPELETNQKEVTASCDLS-----CIVKREK 583
 Db 2064 KF-----EGKCSGPK-----SRNHSKOECCALKGEWGDPCELCPTPEP 2103
 QY 584 RLKKAIRTLURKAVHREQHLQL-----SGKNLDVAKKPP-----RTSER 622
 Db 2104 -----DEAFRLCPGSGTIIVGPDSDAVMDCKEPPDVRHCCQICINTDGS 2148
 QY 623 QAESCVGGGHAENOCV---SRAQTYDGAERILCPNGTFQNEBQMTCPCPRFGN 679
 Db 2149 YRCECPFGVILEGNECVDTDECSVGN-----PCNGTCKRVIGGFEC-TCCEGFE 2197
 QY 680 SGALKTPKAWNMBEGG---LCOPGEYSADGAPCOLCALGTFQPEAGRTSC---FPCGG 733
 Db 2198 PGMPTTCE--DINECAQNELLCAFCRVNTYGSYECK-CPVGVLRE-DRRMCKKDEDECAE 2253
 QY 734 GLATKHQGATSDQGETRYQ-----CSPGHFYNTTTHRCIRCPVGTGYPEGKNKVCSCP 758
 Db 2254 G---KHDCTEKQMECKNLGTWYCIGPGYQVRPDGEGCI-----DENECQTKP 2299
 QY 789 GNTTDFDOSTNITQCKNRGCGELGDT-----GYIESFVPGNYPANETCTWTINPPP 843
 Db 2300 G-----ICENGRCNLTLGTYTCENGDFGTA-----PTQDECL----- 2332
 QY 844 KKRILIVPEILPIDDGCVLVMKTSSTSVTYETCYQYER-----PIATTS 994
 Db 2333 DNRGCVCFSEVF---EMMC-----QIGSSNRNPVTKSECCVGGROWGLHCEICFEGTV 2384
 QY 895 RGSKLMIQFKSNEGNGARGFQVPVYTYDYEDYOLIEDVVRG 936
 Db 2365 AYKLC-----PHRGFMINGADVDE--CKVIHDCVNG 2416
 RESULT 7
 MTN2_HUMAN STANDARD; PRT: 956 AA.
 ID MTN2_HUMAN
 AC Q00339; Q9NSZ1;
 DT 30-MAY-2000 (Rel. 39, Last created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Matrilin-2 precursor.
 GN MATN2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS LONG AND SHORT).
 RX PubMed=11124542;
 RA Muratoglu S., Krysan K., Balazs M., Sheng H., Zakany R., Modis L.,
 RA Kiss I., Deak F.
 RT "Primary structure of human matrilin-2, chromosome location of the
 RT MATN2 gene and conservation of an AT-AC intron in matrilin genes."
 RL Cytogenet. Cell Genet. 90:323-327(2000).
 RN [2]
 RP SEQUENCE OF 644-956 FROM N.A.
 RX MEDLINE=97238863; PubMed=9083061;
 RA Deak F., Piecha D., Bachrati C., Paulsson M., Kiss I.;
 RA "Primary structure and expression of matrilin-2, the closest relative
 RT of cartilage matrix protein within the von Willebrand factor type A-

RT like module superfamily".
 RL J. Biol. Chem. 272:9268-9274(1997).
 RN [3]
 RP SEQUENCE OF 244-956 FROM N.A.
 RC TISSUE=Testis;
 RA Diesterhoef A., Lauber C., Mewes H.-W., Gassenhuber J., Wiemann S.;
 RL Submitted (JAN-2000) to the EMBL/GenBank/CCD databases.
 CC -!- FUNCTION: INVOLVED IN MATRIX ASSEMBLY (BY SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=A: alternative splicing; Named isoforms=2;
 CC Name=Long;
 CC isoId=000339-1; Sequence=Displayed;
 CC Name=Short;
 CC isoId=000339-2; Sequence=VSP_001399;
 CC -!- SIMILARITY: Contains 10 EGF-like domains.
 CC -!- SIMILARITY: Contains 2 VWFA domains.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 EMBL: U69263; AAC5260.2;
 EMBL: AJ27638; CAB70953.1; ALT_INIT.
 HSP: PC1132; IEGF.
 X-REF: 602108;
 GC: GC005578; C:extracellular matrix; NAS.
 InterPro: IPR00152; Asx hydroxyl.
 InterPro: IPR01881; EGF_Ca.
 InterPro: IPR036209; EGF-like.
 InterPro: IPR02035; VWFA_A.
 Pfam: PF00308; EGF; 10.
 Pfam: PF00392; Vwa; 2.
 PRINTS: PR00453; VWFA_D0XAIN.
 SMART: SM00379; EGF_CA; 1.
 SMART: SM00327; VWA; 2.
 PROSITE: PS00303; ASX_HYDROXYL; 9.
 PROSITE: PS01186; EGF_2; 9.
 PROSITE: PS02314; VWFA; 2.
 EGF-like domain; Signal; Glycoprotein; Repeat; Coiled coil;
 Alternative splicing:
 FT SIGNAL 1 POTENTIAL.
 FT CHAIN 24 956 MATRILIN-2.
 FT DOMAIN 57 232 VWFA 1.
 FT DOMAIN 238 278 EGF-LIKE 1.
 FT DOMAIN 279 319 EGF-LIKE 2.
 FT DOMAIN 320 360 EGF-LIKE 3.
 FT DOMAIN 361 401 EGF-LIKE 4.
 FT DOMAIN 402 442 EGF-LIKE 5.
 FT DOMAIN 443 483 EGF-LIKE 6.
 FT DOMAIN 484 524 EGF-LIKE 7.
 FT DOMAIN 525 565 EGF-LIKE 8.
 FT DOMAIN 566 606 EGF-LIKE 9.
 FT DOMAIN 607 647 EGF-LIKE 10.
 FT DOMAIN 655 830 VWFA 2.
 FT DOMAIN 917 955 COILED COIL (POTENTIAL).
 FT DISULFID 242 253 BY SIMILARITY.
 FT DISULFID 249 262 BY SIMILARITY.
 FT DISULFID 264 277 BY SIMILARITY.
 FT DISULFID 283 294 BY SIMILARITY.
 FT DISULFID 293 303 BY SIMILARITY.
 FT DISULFID 305 318 BY SIMILARITY.
 FT DISULFID 324 335 BY SIMILARITY.
 FT DISULFID 331 344 BY SIMILARITY.
 FT DISULFID 346 359 BY SIMILARITY.
 FT DISULFID 365 376 BY SIMILARITY.
 FT DISULFID 372 385 BY SIMILARITY.
 FT DISULFID 387 400 BY SIMILARITY.

```

FT DISULFID 406 417 BY SIMILARITY.
FT DISULFID 413 426 BY SIMILARITY.
FT DISULFID 428 441 BY SIMILARITY.
FT DISULFID 447 458 BY SIMILARITY.
FT DISULFID 454 467 BY SIMILARITY.
FT DISULFID 469 482 BY SIMILARITY.
FT DISULFID 488 499 BY SIMILARITY.
FT DISULFID 495 508 BY SIMILARITY.
FT DISULFID 510 523 BY SIMILARITY.
FT DISULFID 529 540 BY SIMILARITY.
FT DISULFID 536 549 BY SIMILARITY.
FT DISULFID 551 564 BY SIMILARITY.
FT DISULFID 570 581 BY SIMILARITY.
FT DISULFID 577 590 BY SIMILARITY.
FT DISULFID 592 605 BY SIMILARITY.
FT DISULFID 611 622 BY SIMILARITY.
FT DISULFID 618 631 BY SIMILARITY.
FT DISULFID 633 646 BY SIMILARITY.
FT CARBOHYD 221 231 N-LINKED (GLUCNA. . .) (POTENTIAL).
FT CARBOHYD 890 890 N-LINKED (GLUCNA. . .) (POTENTIAL).
FT VARSPLIC 861 879 Missing (in isoform Short).
FT CONFLICT 356 356 K -> E (IN REF. 3).
FT CONFLICT 594 594 V -> E (IN REF. 3).
FT CONFLICT 644 644 R -> G (IN REF. 2).
FT CONFLICT 755 755 F -> L (IN REF. 3).
FT CONFLICT 935 935 L -> F (IN REF. 2).
SQ SEQUENCE 956 AA; 106840 MW; 826B7F347178FC80 CRC64;

Query Match 12.2%; Score 684.5; DB 1; Length 956;
Best Local Similarity 32.3%; Pred. No. 48-38;
Matches 146; Conservative 61; Mismatches 186; Indels 59; Gaps 13;

QY 62 QNTPTSYKCSKPGY--QGEGRQCEIDECGNELNGCVHDCLNIPGNYRCTCFDGEWL 119
Db 253 CINIPGSYVCRCKQYILNSDTCRIQDLCAE-DHNCCEQLCVNPGSFVQCYSYVAL 311
QY 120 AUDGHNCLDVBCELENNGGCQHTCVNMGSYSCCKKEGFFLSDNQHTCIHREGLSCWN 179
Db 312 AEDGKRCVAVDYCASENHGCHECVNADSGYLQCCHGFPALNPDKTKTID----YCAS 367
QY 180 KDHGSHICEAPRGSVACECPGPELAKNORDC--ILTCNHNGGCGHSCDDTADGPEC 237
Db 368 SHHQOHECVNTD-DYSCHCLKLGFTLPDPKTKRINVCALNKPGCEHCNVMEESYYC 426
QY 238 SCHPOYMHITDGRSCLEREDTVLEVTESNTTSVVDGKFKVRRLMETCAVNGGCDRTC 297
Db 427 RCHRGYTLDPNGKTCR-----VDHCAQODHGCQGLC 458
QY 298 KOTSTGVHSCPVGFTLQDGTCKDIDECOTRNGGCDHFCKN-VGSPDGGCKKGFKLLT 357
Db 459 LNTEDSFVQCSEGFLLNEDLTKTSRDVYCLSDHGCYSVNMMDRSFACQCPGHLVRS 518
QY 358 DEKSCQDVBECSL-DRTCDHSCINHPGTFACACNRYTLY--GFTHCGDTNCSINNGGC 414
Db 519 DGKTCALSKALGDMGCHSCSVSSDSFVQCQFEGYLLREDKTK-CRKQVQCQIDHGC 577
QY 415 QQCVNTVGSYEQCHPGYKLMNKKDCVEVGLLPTSVSPRSLHCGKS-----GGGQCC 470
Db 578 EHICVNSDDSYTCECLVGFRLAEDGKRC-RRKDVCK-----STHGCEHCIVNGKSY 629
QY 471 FLRCHSGIHLSDVTTIRTSVTFKLNKGKCSL 502
Db 630 ICKSEGFVLAEDGRCK-----KCTEGPIDL 656

RESULT 8
ID_FBN1_PIG STANDARD; PRT; 2871 AA.
AC Q9TV36;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)

```

```

DE GN Fibrillin 1 precursor.
OS Sus scrofa (Pig).
OC Eukaryota; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.
OX NCBI_TaxID=9923;
RN [1]
RP SEQUENCE FROM N.A.
RC T-SSU=Lung;
RX MEDLINE=99156838; PubMed=10036187;
RA Biery N.J.; Eldadah Z.A.; Moore C.S.; Stetten G.; Spencer F.;
RA Dietz H.C.;
RT "Revised genomic organization of FBN1 and significance for regulated
RZ gene expression.";
RZ Genomics 56:70-77(1999).
CC -!- FUNCTION: STRUCTURAL COMPONENT OF CONNECTIVE TISSUE MICROFIBRILS
CC THAT BINDS CALCIUM. FIBRILLIN-1-CONTAINING MICROFIBRILS PROVIDE
CC LONG-TERM FORCE BEARING STRUCTURAL SUPPORT.
CC -!- PTM: FORMS INTERMOLECULAR DISULFIDE BONDS EITHER WITH OTHER
CC MICROFIBRILS.
CC -!- SIMILARITY: Contains 47 EGF-like domains.
CC -!- SIMILARITY: Contains 7 TGF-beta binding protein (TGFBP) domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EXBL: AF073800; AAD50328.1; .
DR HSPB; P35555; IAPC.
DR InterPro: IPR000152; Asx hydroxyl.
DR InterPro: IPR001881; EGF-Ca.
DR InterPro: IPR006209; EGF-like.
DR InterPro: IPR006212; Fibril-associ.
DR Pfam: PF00008; EGF; 45.
DR Pfam: PF00683; TB; 9.
DR SMART: SM00179; EGF CA; 40.
DR PROSITE: PS00010; ASX HYDROXYL; 41.
DR PROSITE: PS00022; EGF_1; 2.
DR PROSITE: PS01186; EGF_2; 36.
DR PROSITE: PS01187; EGF CA; 41.
DR Extracellular matrix; Calcium-binding; Glycoprotein; EGF-like domain;
KW Repeat; Signal; Multigene family.
FT SIGNAL 27 POTENTIAL.
FT CHAIN 28 2871 FIBRILLIN 1.
FT DOMAIN 81 112 EGF-LIKE 1.
FT DOMAIN 115 146 EGF-LIKE 2.
FT DOMAIN 147 178 EGF-LIKE 3.
FT DOMAIN 246 287 EGF-LIKE 4. CALCIUM-BINDING.
FT DOMAIN 288 329 EGF-LIKE 5. CALCIUM-BINDING.
FT REPEAT 330 390 TGFBP 1.
FT DOMAIN 392 446 PRO-RICH.
FT DOMAIN 449 489 EGF-LIKE 6.
FT DOMAIN 490 529 EGF-LIKE 7. CALCIUM-BINDING.
FT DOMAIN 530 571 EGF-LIKE 8. CALCIUM-BINDING.
FT DOMAIN 572 613 EGF-LIKE 9. CALCIUM-BINDING.
FT DOMAIN 613 653 EGF-LIKE 10. CALCIUM-BINDING.
FT REPEAT 654 722 TGFBP 2.
FT DOMAIN 723 764 EGF-LIKE 11. CALCIUM-BINDING.
FT DOMAIN 765 806 EGF-LIKE 12. CALCIUM-BINDING.
FT DOMAIN 807 846 EGF-LIKE 13. CALCIUM-BINDING.
FT DOMAIN 849 910 EGF-LIKE 14. CALCIUM-BINDING.
FT REPEAT 952 1027 TGFBP 3.
FT DOMAIN 1028 1069 EGF-LIKE 15. CALCIUM-BINDING.
FT DOMAIN 1070 1112 EGF-LIKE 16. CALCIUM-BINDING.
FT DOMAIN 1113 1154 EGF-LIKE 17. CALCIUM-BINDING.
FT DOMAIN 1155 1197 EGF-LIKE 18. CALCIUM-BINDING.
FT DOMAIN 1197 1237 EGF-LIKE 19. CALCIUM-BINDING.
FT DOMAIN 1238 1279 EGF-LIKE 20. CALCIUM-BINDING.

```

```

FT DOMAIN 1280 EGF-LIKE 21, CALCIUM-BINDING.
FT DOMAIN 1322 EGF-LIKE 22, CALCIUM-BINDING.
FT DOMAIN 1363 EGF-LIKE 23, CALCIUM-BINDING.
FT DOMAIN 1404 EGF-LIKE 24, CALCIUM-BINDING.
FT DOMAIN 1446 EGF-LIKE 25, CALCIUM-BINDING.
FT DOMAIN 1487 EGF-LIKE 26, CALCIUM-BINDING.
FT REPEAT 1528 TGFBP 4.
FT DOMAIN 1606 EGF-LIKE 27, CALCIUM-BINDING.
FT DOMAIN 1648 EGF-LIKE 28, CALCIUM-BINDING.
FT REPEAT 1689 TGFBP 5.
FT DOMAIN 1766 EGF-LIKE 29, CALCIUM-BINDING.
FT DOMAIN 1808 EGF-LIKE 30, CALCIUM-BINDING.
FT DOMAIN 1849 EGF-LIKE 31, CALCIUM-BINDING.
FT DOMAIN 1891 EGF-LIKE 32, CALCIUM-BINDING.
FT DOMAIN 1930 EGF-LIKE 33, CALCIUM-BINDING.
FT DOMAIN 1973 EGF-LIKE 34, CALCIUM-BINDING.
FT DOMAIN 2013 EGF-LIKE 35, CALCIUM-BINDING.
FT REPEAT 2055 TGFBP 6.
FT DOMAIN 2127 EGF-LIKE 36, CALCIUM-BINDING.
FT DOMAIN 2166 EGF-LIKE 37, CALCIUM-BINDING.
FT DOMAIN 2206 EGF-LIKE 38, CALCIUM-BINDING.
FT DOMAIN 2247 EGF-LIKE 39, CALCIUM-BINDING.
FT DOMAIN 2291 EGF-LIKE 40, CALCIUM-BINDING.
FT REPEAT 2333 TGFBP 7.
FT DOMAIN 2402 EGF-LIKE 41, CALCIUM-BINDING.
FT DOMAIN 2444 EGF-LIKE 42, CALCIUM-BINDING.
FT DOMAIN 2485 EGF-LIKE 43, CALCIUM-BINDING.
FT DOMAIN 2524 EGF-LIKE 44, CALCIUM-BINDING.
FT DOMAIN 2567 EGF-LIKE 45, CALCIUM-BINDING.
FT DOMAIN 2607 EGF-LIKE 46, CALCIUM-BINDING.
FT DOMAIN 2648 EGF-LIKE 47, CALCIUM-BINDING.
FT DISULFID 85 BY SIMILARITY.
FT DISULFID 89 BY SIMILARITY.
FT DISULFID 102 BY SIMILARITY.
FT DISULFID 119 BY SIMILARITY.
FT DISULFID 123 BY SIMILARITY.
FT DISULFID 136 BY SIMILARITY.
FT DISULFID 150 BY SIMILARITY.
FT DISULFID 154 BY SIMILARITY.
FT DISULFID 168 BY SIMILARITY.
FT DISULFID 250 BY SIMILARITY.
FT DISULFID 257 BY SIMILARITY.
FT DISULFID 273 BY SIMILARITY.
FT DISULFID 292 BY SIMILARITY.
FT DISULFID 299 BY SIMILARITY.
FT DISULFID 313 BY SIMILARITY.
FT DISULFID 453 BY SIMILARITY.
FT DISULFID 460 BY SIMILARITY.
FT DISULFID 476 BY SIMILARITY.
FT DISULFID 494 BY SIMILARITY.
FT DISULFID 504 BY SIMILARITY.
FT DISULFID 513 BY SIMILARITY.
FT DISULFID 515 BY SIMILARITY.
FT DISULFID 534 BY SIMILARITY.
FT DISULFID 541 BY SIMILARITY.
FT DISULFID 557 BY SIMILARITY.
FT DISULFID 576 BY SIMILARITY.
FT DISULFID 582 BY SIMILARITY.
FT DISULFID 598 BY SIMILARITY.
FT DISULFID 617 BY SIMILARITY.
FT DISULFID 623 BY SIMILARITY.
FT DISULFID 639 BY SIMILARITY.
FT DISULFID 727 BY SIMILARITY.
FT DISULFID 734 BY SIMILARITY.
FT DISULFID 750 BY SIMILARITY.
FT DISULFID 769 BY SIMILARITY.
FT DISULFID 776 BY SIMILARITY.
FT DISULFID 792 BY SIMILARITY.
FT DISULFID 811 BY SIMILARITY.
FT DISULFID 816 BY SIMILARITY.
FT DISULFID 832 BY SIMILARITY.
FT DISULFID 914 BY SIMILARITY.
FT DISULFID 921 BY SIMILARITY.
FT DISULFID 937 BY SIMILARITY.

```

```

FT DISULFID 1032 BY SIMILARITY.
FT DISULFID 1039 BY SIMILARITY.
FT DISULFID 1055 BY SIMILARITY.
FT DISULFID 1074 BY SIMILARITY.
FT DISULFID 1081 BY SIMILARITY.
FT DISULFID 1095 BY SIMILARITY.
FT DISULFID 1111 BY SIMILARITY.
FT DISULFID 1117 BY SIMILARITY.
FT DISULFID 1124 BY SIMILARITY.
FT DISULFID 1159 BY SIMILARITY.
FT DISULFID 1171 BY SIMILARITY.
FT DISULFID 1201 BY SIMILARITY.
FT DISULFID 1221 BY SIMILARITY.
FT DISULFID 1238 BY SIMILARITY.
FT DISULFID 1254 BY SIMILARITY.
FT DISULFID 1263 BY SIMILARITY.
FT DISULFID 1265 BY SIMILARITY.
FT DISULFID 1278 BY SIMILARITY.
FT DISULFID 1284 BY SIMILARITY.
FT DISULFID 1291 BY SIMILARITY.
FT DISULFID 1307 BY SIMILARITY.
FT DISULFID 1326 BY SIMILARITY.
FT DISULFID 1333 BY SIMILARITY.
FT DISULFID 1350 BY SIMILARITY.
FT DISULFID 1367 BY SIMILARITY.
FT DISULFID 1374 BY SIMILARITY.
FT DISULFID 1391 BY SIMILARITY.
FT DISULFID 1408 BY SIMILARITY.
FT DISULFID 1415 BY SIMILARITY.
FT DISULFID 1461 BY SIMILARITY.
FT DISULFID 1456 BY SIMILARITY.
FT DISULFID 1472 BY SIMILARITY.
FT DISULFID 1491 BY SIMILARITY.
FT DISULFID 1497 BY SIMILARITY.
FT DISULFID 1513 BY SIMILARITY.
FT DISULFID 1610 BY SIMILARITY.
FT DISULFID 1617 BY SIMILARITY.
FT DISULFID 1633 BY SIMILARITY.
FT DISULFID 1652 BY SIMILARITY.
FT DISULFID 1658 BY SIMILARITY.
FT DISULFID 1674 BY SIMILARITY.
FT DISULFID 1770 BY SIMILARITY.
FT DISULFID 1777 BY SIMILARITY.
FT DISULFID 1793 BY SIMILARITY.
FT DISULFID 1812 BY SIMILARITY.
FT DISULFID 1824 BY SIMILARITY.
FT DISULFID 1835 BY SIMILARITY.
FT DISULFID 1853 BY SIMILARITY.
FT DISULFID 1860 BY SIMILARITY.
FT DISULFID 1876 BY SIMILARITY.
FT DISULFID 1895 BY SIMILARITY.

Query Match 12.1%; Score 678; DB 1; length 287;
Best Local Similarity 22.7%; Pred. No. 3,7e-37;
Matches 270; Conservative 109; Mismatches 377; Indels 434; Gaps 51;

Qy 45 DVDECAQGLDCHADALCCQNTFTSYKCSCKPGYQGEGRCEDECGNELN----- 95
Db 1363 DLDECSNGTHMCSQHACDKNTMGSYRLCKEYTGDFTCALDSCSNVKLGKGVQCILY 1422

Qy 96 --GG--CVHD-----CLNIPGNVRCCTCGFMALHDG 123
Db 1423 APOGYHCEYDMGVFVSADRSKSCVDSDECSJPNICVFGTCHNLPLGRCECEIGYELDRSG 1492

Qy 124 HNCILDVCELENNGGQHTCVNVMGSYECCCKEYGFPLSDNHTCTH-----RS 171
Db 1463 GNCIDVNECLEPPTC:SGNCVNTPGSYTCVCPDPDELNPTRVGCVDTRSGNCYLVRPRG 1542

Qy 172 EGG-LSCMN-----KDHGCSHI----- 187
Db 1543 DNGDTACSNIEIGVGVSKASCCSLGKAWCTPCEQCPVNTSEYKILCPGGEGFRENPIV 1602

Qy 188 -----CKEAPR-----GSVACEPRPGFELAKNQDCC---TCNH-----GNGG 223
Db 1603 ILEDIDECCELFGLCOGKGCINTFGSQCRCTGYLYNEDTRVCDVNECECTPGICGPGS- 1661

```


FT CHAIN 21 1394 LATENT TRANSFORMING GROWTH FACTOR BETA
 FT DOMAIN 300 340 BINDING PROTEIN, ISOFORM 1S.
 FT REPEAT 348 412 EGF-LIKE 1, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 546 587 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 588 629 EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 630 670 EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 671 710 EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 711 751 EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 752 792 EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 793 833 EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 834 874 EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 875 916 EGF-LIKE 10, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 917 958 EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 959 1001 EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
 FT REPEAT 1017 1084 REPEAT B.
 FT DOMAIN 1097 1139 EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
 FT REPEAT 1190 1262 REPEAT C.
 FT DOMAIN 1191 1280 EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 1294 1334 EGF-LIKE 15.
 FT DOMAIN 1335 1379 EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
 FT SITE 947 949 CELL ATTACHMENT SITE (POTENTIAL).
 FT DISULFID 304 315 BY SIMILARITY.
 FT DISULFID 310 324 BY SIMILARITY.
 FT DISULFID 326 339 BY SIMILARITY.
 FT DISULFID 550 562 BY SIMILARITY.
 FT DISULFID 557 571 BY SIMILARITY.
 FT DISULFID 573 586 BY SIMILARITY.
 FT DISULFID 592 604 BY SIMILARITY.
 FT DISULFID 599 613 BY SIMILARITY.
 FT DISULFID 615 628 BY SIMILARITY.
 FT DISULFID 634 645 BY SIMILARITY.
 FT DISULFID 640 654 BY SIMILARITY.
 FT DISULFID 657 669 BY SIMILARITY.
 FT DISULFID 675 686 BY SIMILARITY.
 FT DISULFID 681 695 BY SIMILARITY.
 FT DISULFID 698 709 BY SIMILARITY.
 FT DISULFID 715 726 BY SIMILARITY.
 FT DISULFID 721 735 BY SIMILARITY.
 FT DISULFID 737 750 BY SIMILARITY.
 FT DISULFID 756 767 BY SIMILARITY.
 FT DISULFID 762 776 BY SIMILARITY.
 FT DISULFID 778 791 BY SIMILARITY.
 FT DISULFID 797 808 BY SIMILARITY.
 FT DISULFID 803 817 BY SIMILARITY.
 FT DISULFID 819 832 BY SIMILARITY.
 FT DISULFID 838 850 BY SIMILARITY.
 FT DISULFID 845 859 BY SIMILARITY.
 FT DISULFID 861 873 BY SIMILARITY.
 FT DISULFID 879 891 BY SIMILARITY.
 FT DISULFID 895 900 BY SIMILARITY.
 FT DISULFID 902 915 BY SIMILARITY.
 FT DISULFID 921 933 BY SIMILARITY.
 FT DISULFID 927 942 BY SIMILARITY.
 FT DISULFID 944 957 BY SIMILARITY.
 FT DISULFID 957 970 BY SIMILARITY.
 FT DISULFID 970 984 BY SIMILARITY.
 FT DISULFID 986 1000 BY SIMILARITY.
 FT DISULFID 1101 1114 BY SIMILARITY.
 FT DISULFID 1109 1123 BY SIMILARITY.
 FT DISULFID 1125 1138 BY SIMILARITY.
 FT DISULFID 1144 1155 BY SIMILARITY.
 FT DISULFID 1150 1164 BY SIMILARITY.
 FT DISULFID 1166 1179 BY SIMILARITY.
 FT DISULFID 1298 1309 BY SIMILARITY.
 FT DISULFID 1304 1318 BY SIMILARITY.
 FT DISULFID 1320 1333 BY SIMILARITY.
 FT DISULFID 1339 1354 BY SIMILARITY.
 FT DISULFID 1349 1363 BY SIMILARITY.
 FT DISULFID 1365 1378 BY SIMILARITY.
 FT MOD_RES 647 647 HYDROXYLATION.
 FT MOD_RES 810 810 HYDROXYLATION.
 FT CARBOHYD 21 21 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 52 52 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 98 98 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 294 294 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 870 870 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 923 923 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1039 1039 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT FTID-CAR 000184
 FT CONFLICT 365 365 Y -> H (IN REF. 2).
 FT CONFLICT 504 504 A -> T (IN REF. 2).
 FT CONFLICT 1334 1334 F -> V (IN REF. 2).
 SC SEQUENCE 1394 AA: 152791 MW: 27581.440327C1 CRC64;
 Query Match 11.1%; Score 624.5; DB: Length 1394;
 Best Local Similarity 24.9%; Pred. No. 66-34;
 Matches 240; Conservative 94; Mismatches 334; Indels 297; Gaps 48;
 QY 45 DVDECAQGLDDCHADALCALQNTPTSYKSCPKPVQ--GEGROCEDIDECGNELNGSCVHD- 101
 DB 546 EINECTVNPICGA-GHCINLPVRYTCICEGYRSEQQKCVDIDEC-TQVCHLCSQGR 603
 QY 102 CLNIPGNRYCTCPDGMJAHDHNCILDVDECLNNGSGCHTCTVYMGSEYEC-CKEKGFL 160
 DB 604 CENTEGSFLLCPAGPMASSEGTNCIDVDSCLRPDYCGEGHCNTVGAFCRCYCDSGVM 663
 QY 161 SDNQHTCIHREBESLSCKMNDHGCISHCKEAPRGSVACRCFOFELAKNQRCILTCNHG 220
 DB 664 TQR-----GRCEIDELNPSTCPDEQVNSP-GSYQC-----VPCTRG 701
 QY 221 NGCOHSCDUTADGPESCSCHPQKMTDGRSCJEREDTVLEVETSTNTTSVWZGDKEVRR 260
 DB 702 FRGNQQLDVE-----CUE----- 717
 QY 281 LLMETCAVANGSCDRCKDTSTGVHSCVPVGTLLDQKTKGIDECQTRNGGCHFFCKY 340
 DB 718 --PNVCA--NG-----DCSNLESGYMSCHKVTRTPDKHCHREDIDECQGNLVNGGCKN 769
 QY 341 IVGSPGCGKKGFKLLTDEKSKQVDECSLDRTCH-SCIRHEGTFAACNRGVTLVGF 398
 DB 770 TEGSFCTCGGVQVLSAARKQCEIDECOKHLCARGQCRNTEGSCFQVCDQGYRAGSG 829
 QY 399 THCGDNESINNGGCGQ--VCVNTVSGSEYEQCHPGYKJHWNK-----KDCVEVKGLPTS 452
 DB 830 CHCEDINELEKSVQQRGDCNTAGSYDCTCPDGFQDKNKTCQDINEC-EHPGJ----- 584
 QY 453 VSPRVSJHCKSG--CGDGCF-LRCHSGIHLSSVTTIRTSTVTKJREGKSLXNAELF 508
 DB 885 -----CPOGECUNTFGSHCVCCQGSFISADQRT-----CRDI 918
 QY 509 PEGLRPALPEKH---SSVKESFRYVNLTCSSGKV?-----GAPGRP-- 547
 DB 919 DECVNNTVCDSHGFCDNTAGSFRCL--CYQGFQAFQDQGGCVNNECELLSGVCGEAF 975
 QY 549 STPKEMFIV-----EPELEINQKVTASCDLSCLVIRKTEKJLRKAINTLRKA----- 595
 DB 976 ENVEGSLVCADENQSEYSPVTCQCRSTSTLDVDVDQFKEEKKECYVNLNDASLCNV 1035
 QY 596 ---VHREQFHLCL---GMNLDVAKKPPRTSERQAESCGVGQH-----AEN- 636
 DB 1036 LAENVTKQECCTSGAGWGDNCEIFPCPVLTGTAETEMCPKCKGFPVPAESSEAGENY 1095
 QY 637 ---QCV-----SCRAGTYVDGARECI-----LCPNGTF 662
 DB 1096 KDADECLLFGQIECKNGFCLNTRPGVEYCKQGYVDPVKLQCFDMECDQSPSSCIDGQC 1155
 QY 663 QNEEGQWTC-----EPCRPNGS-----G 681
 DB 1156 VNTGYSNCFCTHPVLDASEKRCIRPASENCEIBETDVQDLCHLHLSDEYVVCRLVW 1215
 QY 682 ALXT-----PEAWNSECGGLCPGGEYSADGAPCOLCALGTGFQPEAGRTSCF----- 729
 DB 1216 KQTTYECCLLYGEAWGM-QC-ALC-PLKDSDDYACLNIPIUTGRQP-YGRDALVDSE 1271
 QY 730 ---PCGGGLATKIQGATSFQDCETRVCQSPGHFYNTTHRCIRCPVG-----TYQPEF 779


```

Db      1272 QVTEADPYFIQDRFLNFEELQAE-EC--GILNGCNGRCVRVQEGYTCDCCLDGYHLST 1328
QY      780 GKQNCVSCPGNTTDPDGSTN-ITQCKNRCCGELGDF-----TGVIES--PNY--PGNY 829
Db      1329 AKWTCTFD-----VNECDLNNRMSLCKNAKINTDGSYKCLPGVSDPKPNYCTP-NT 1383
QY      830 PANTE 834
Db      1384 ALNLE 1388

RESULT 10
LTBL_HUMAN
ID LTBL_HUMAN STANDARD; PRT: 1595 AA.
AC Q14766;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Latent transforming growth factor beta binding protein, isoform 1L
DE Precursor (LTBP-1) (transforming growth factor beta-1 binding protein
DE 1) (TGF-beta1-BP-1).
GN LTBP1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE OF 1-346 FROM N.A.
RC TISSUE=Blood;
RX MEDLINE=96125117; PubMed=8537398;
RA Olofsson A., Ichijo H., Moren A., ten Dijke P., Miyazono K.,
RA Heldin C.-H.;
RT "Efficient association of an amino-terminally extended form of human
RT latent transforming growth factor-beta binding protein with the
RT extracellular matrix";
RL J. Biol. Chem. 270:31294-31297(1995).
RN [2]
RP SEQUENCE OF 347-1595 FROM N.A.
RC TISSUE=Fibroblast, and Platelet;
RX MEDLINE=90275601; PubMed=2350783;
RA Kanzaki T., Olofsson A., Moren A., Wernstedt C., Hellman U.,
RA Miyazono K., Claesson-Welsh L., Heldin C.-H.;
RT "TGF-beta 1 binding protein: a component of the large latent complex
RT of TGF-beta 1 with multiple repeat sequences.";
RL Cell 61:1051-1061(1990).
RN [3]
RP INTERACTION WITH FIBRILLIN.
RX PubMed=12429738;
RA Isogai Z., Ono R.N., Ushiro S., Keene D.R., Chen Y., Mazzieri R.,
RA Charbonneau N.L., Reinhardt D.P., Rifkin D.B., Sakai L.Y.;
RT "Latent transforming growth factor beta-binding protein 1 interacts
RT with fibrillin and is a microfibril-associated protein.";
RL J. Biol. Chem. 278:2750-2757(2003).
CC -/- SUBUNIT: The large latent complex of TGF-beta1 from platelets is
CC composed of the TGF-beta1 molecule noncovalently associated with a
CC disulfide-bonded complex of a dimer of the N-terminal propeptide
CC of the TGF-beta1 precursor and a third component denoted TGF-
CC beta1-BP. TGF-beta1-BP does not bind directly to active TGF-BETA1.
CC Binds to fibrillin.
CC -/- SUBCELLULAR LOCATION: Secreted.
CC -/- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Long;
CC IsoId=Q14766-1; Sequence=Displayed;
CC Name=Short;
CC IsoId=P22064-1; Sequence=External;
CC -/- TISSUE SPECIFICITY: The long isoform is found in fibroblasts.
CC -/- PTM: CONTAINS HYDROXYLATED ASPARAGINE RESIDUES (BY SIMILARITY).
CC -/- SIMILARITY: Contains 16 EGF-like domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

```

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: J48925; AAA96327.1;
CC EMBL: M34057; AAA6160.1; ALT_INIT.
CC HSSP: PC8739; BPF9.
CC MIN: J5C39C;
CC GC: GC005578; C:extracellular matrix; NAS.
CC GC: GC0305024; F:transforming growth factor-beta receptor ac...; NAS.
CC InterPro: IPR000152; Asx_hydroxyl.
CC InterPro: IPR001891; EGF_Ca.
CC InterPro: IPR036203; EGF_Like.
CC InterPro: IPR002112; Fibrin_assoc.
CC Pfam: PF03006; EGF_15.
CC Pfam: PF03683; TB_4.
CC SMART: SMC379; EGF_CA_13.
CC PROSITE: PS00013; ASX_HYDROXYL; 13.
CC PROSITE: PS00022; EGF_1; 2.
CC PROSITE: PS01184; EGF_2; 11.
CC PROSITE: PS01187; EGF_CA_15.
CC Growth factor binding; Repeat; EGF-like domain; Hydroxylation; Signal;
KW Glycoprotein; Alternative splicing.
FT SIGNAL 1 23
FT CHAIN 24 1595
FT DOMAIN 501 541
FT REPEAT 549 613
FT DOMAIN 747 788
FT DOMAIN 789 830
FT DOMAIN 831 871
FT DOMAIN 872 911
FT DOMAIN 912 952
FT DOMAIN 953 993
FT DOMAIN 994 1034
FT DOMAIN 1035 1075
FT DOMAIN 1076 1117
FT DOMAIN 1118 1159
FT DOMAIN 1160 1202
FT REPEAT 1218 1285
FT DOMAIN 1298 1340
FT REPEAT 1391 1463
FT DOMAIN 1495 1535
FT DOMAIN 1536 1580
FT SITE 1648 1050
FT DISULF 505 516
FT DISULF 511 525
FT DISULF 527 540
FT DISULF 751 763
FT DISULF 758 772
FT DISULF 774 787
FT DISULF 793 805
FT DISULF 806 814
FT DISULF 816 829
FT DISULF 835 846
FT DISULF 841 855
FT DISULF 858 870
FT DISULF 876 887
FT DISULF 882 896
FT DISULF 899 910
FT DISULF 916 927
FT DISULF 922 936
FT DISULF 938 951
FT DISULF 957 969
FT DISULF 963 977
FT DISULF 979 992
FT DISULF 998 1009
FT DISULF 1004 1018
FT DISULF 1020 1033
FT DISULF 1039 1051

```

Query Match 11.1%; Score 624.5; DB 1; Length 1595;
 Best Local Similarity 24.9%; Pred. No. 7e-34;
 Matches 240; Conservative 94; Mismatches 334; Indels 297; Gaps 48;

QY 45 DVDECAQGLDDHADALQNTTYSKSCPKPGY--GEGRCQEDIDECNELNGCVHD- 101
 DB 747 EINECTVNPIDCGA-GHCINLPVYTICVGYRSEQQRKQVD-DEC-TQVQH-CSQGR 304
 QY 102 CLNIPQNYRCTCFDGMALHDGNCIDVDECLNNGCQOHTCVNMGSECC-CKEGSFL 160
 DB 805 CENTGSEFLCICPAGPMASEEGTNCIDVDECLRPVCGEGHCWNTVGAPRCVCDGSRM 964
 QY 161 SONQHTCIHRSEGLSCMNKDHGCHICKAEPGRSVACRCRPFELAKQQRCDITCNKG 220
 DB 865 TOR-----GRCED:DECLNPSTCPDQCVNSP-GSYQC-----VPTGEG 902
 QY 221 NGCQHSDDTDGPECSCHPQVYMTDGRSLEREDTVLENTSNTSVVDGKVKRR 280
 DB 903 FRWNGQCLDVDE-----CLE----- 919
 QY 281 LLMETCAVNGGCDRTCKDSTGVHSCPCVGFPTQLDGTCKDIDECQCRNSGCDHFCNK 340
 DB 919 --PNVCA--NG-----DCSNLEGS:WCSCHKGYTRTPDHKHCXRDIDECQGNLCVNGQCKN 970
 QY 341 IVGSFDCGCKGFKLLTDEKSCQDVDECLDRDCH-SCINHPGTACACNRGYLYGF- 398
 DB 971 TEGSFRTCTCGYOLSAKQDQEDIDECQHRHLCAHQCRNTEGSPQCVCDGYRASGLG 1030
 QY 399 THCGDITNEGSINNGCQO-VCVNTVGSYECQHPGYKLHWK-----KDCVEKGLLPTS 452
 DB 1031 DHCEDINECLDKSVQCRGDCINTAGSYDCTPDGFLQDNCITQDINBC-EHPGJ----- 1085
 QY 453 VSPRVLHCKSG--GGQGF-LRCHSGTHLSSDVTTITRTSVFKLNEGKSLKNAELF 508
 DB 1086 -----CQPOEC:NTGSEFVCVQCFSTISADGT-----CEDI 1119
 QY 509 PEGLRPALPEKH---SSVKESPRYVNLTCSSGKQVP-----GAPGRP-- 547
 DB 1:20 DECVMNTVCDSHGFCNDNTAGSPFCT---CYQGFAQPCQGGCVVYNECELLSSVGEAF 1176
 QY 548 STPKEMFIV-----EPELETNQKEVNTASCLSCIVKTEKRLKAITLKA----- 595

Db 1177 ENVEGSFLVCACENQESYSPMTQCCSRSTSDLDVDVQDPKEEKKECYNYLNDASLCNV 1236
 QY 596 ----VHREQPHLQLS---GMNLDVAKKPPRTSERQAESCGVQGH-----AEN- 636
 Db 1237 LAPNVTQKECCCTSGAGWGNCEIFPCPVLGTAETEMCPKGFVPAGESSEAGENY 1296
 QY 637 ----QCV-----SCRAGTYVDGARERCI-----LCPNGTF 662
 Db 1297 KDADECLFQGEIKCKGFCNTRPGVECYCKQGTYYDPVKLOCFMDEQDSSSCIDGQC 1356
 QY 663 QNEEGQWTC-----EPCPRGNS-----G 681
 Db 1357 VNTGSSVNCFTHPMVLDAKRCIRPAESNBOIBETDYYQDLCEHLSDEYVCSRPLVS 1416
 QY 682 ALKT-----FEANMSECGGLQCPGEYSADGFAPCQCALGTQPPAGRTSCF---- 729
 Db 1417 KQTTYTECCCLYGEAWG-QC-ALC-PLKDSDDYAQLCNIPVTGRRQP-VGRDALVDFSE 1472
 QY 730 ---PCGGGLATKHGATSFQDCETRVQCSGPHFYNTHRCIRCPVG-----TYQPEF 779
 Db 1473 QYTPEADPYFIQDRFLNSFEELQAE-SC--GLNSCENGRCVRVQEGYTCDCGLGYHLET 1529
 QY 780 GKXNVSCPNNTTCDPDGSTN-ITCKNRCCGEGGDF-----TGYES--PNY--PGHY 829
 Db 1530 AKMTCPD-----VNECEDLNRMSCNAKAC:NTDGSYKCLCLPGVVPDCKNYCTP:NT 1584
 QY 830 PANTE 834
 Db 1585 ALNJE 1599

RESULT 11
 ID: LBS_XCUSE STANDARD: PRT: 1359 AA.
 AC Q8CG18; Q8BNW7; Q8C7F5; Q8C1RC;
 DT 15-SEP-2003 (Rel. 42, Created;
 DT 15-SEP-2003 (Rel. 42, Last sequence update;
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Latent transforming growth factor beta binding protein, isoform 1S
 DE precursor (LTBP-1) (transforming growth factor beta-1 binding protein
 DE 1) (TGF-beta1-SP-1).
 GN LTBP1.
 CS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1].
 RP SEQUENCE FROM N.A.
 RX PubMed=12711388;
 RA Noguera I., Obata H., Gualandris A., Cowin P., Rifkin D.B.;
 RT "Molecular cloning of the mouse Ltbp-1 gene reveals tissue specific
 expression of alternatively spliced forms.";
 RL Gene 308:31-41(2003).
 RN [2].
 RP SEQUENCE FROM N.A.
 RX PubMed=12711389;
 RA Weiskirchen R., Moser M., Guenther K., Weiskirchen S., Gressner A.M.;
 RT "The murine latent transforming growth factor-beta binding protein
 (Ltbp-1) is alternatively spliced, and maps to a region syntenic to
 human chromosome 2p21-22.";
 RL Gene 308:43-52(2003).
 RN [3].
 RP SEQUENCE OF 788-1389 FROM N.A.
 PC STRAIN=C57BL/6J; TISSUE=Aorta, Liver, and Vein;
 RX PubMed=12466851;
 RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
 RA Nakaide T., Osato N., Saito R., Suzuki H., Yamanaka T., Kiyosawa H.,
 RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gotohori T.,
 RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
 RA Schrim D.M., Kanapin A., Matsuda H., Bacalov S., Heise K.W.,
 RA Blake C.A., Bratt D., Bruscia V., Chothia C., Corbani J.E., Cousins S.,
 RA Dalla E., Dragani Z.A., Fletcher C.F., Forrest A., Frazer K.S.,

RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA Kanai A., Kawai H., Kawasawa Y., Kedzierski R.M., King B.L.,
RA Konoaga A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
RA Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,
RA Sultana R., Tanaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
RA Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
RA Yuan Z., Zavalon M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA Hirozane-Kishikawa T., Kono H., Nakamura M., Sakazume N., Sato K.,
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shingawa A.,
RA Yasuishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
RA Birney E., Hayashizaki Y.,
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs";
RL Nature 420:563-573(2002).
CC -:- SUBUNIT: The large latent complex of TGF-beta1 from platelets is
CC composed of the TGF-beta1 molecule noncovalently associated with a
CC disulfide-bonded complex of a dimer of the N-terminal propeptide
CC of the TGF-beta1 precursor and a third component denoted TGF-
CC beta1-BP. TGF-beta1-BP does not bind directly to active TGF-BETA1.
CC Binds to fibrillin (By similarity).
CC -:- SUBCELLULAR LOCATION: Secreted (By similarity).
CC -:- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1S;
CC IsoId=Q8CG18-1; Sequence=Displayed;
CC Name=1L;
CC IsoId=Q8CG19-1; Sequence=External;
CC -:- PTM: Contains hydroxylated asparagine residues (By similarity).
CC -:- PTM: The N-terminus is blocked (By similarity).
CC -:- SIMILARITY: Contains 16 EGF-like domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF346465; AAN77251.1; JOINED.
DR EMBL; AF346438; AAN77251.1; JOINED.
DR EMBL; AF346439; AAN77251.1; JOINED.
DR EMBL; AF346440; AAN77251.1; JOINED.
DR EMBL; AF346441; AAN77251.1; JOINED.
DR EMBL; AF346442; AAN77251.1; JOINED.
DR EMBL; AF346443; AAN77251.1; JOINED.
DR EMBL; AF346444; AAN77251.1; JOINED.
DR EMBL; AF346445; AAN77251.1; JOINED.
DR EMBL; AF346446; AAN77251.1; JOINED.
DR EMBL; AF346447; AAN77251.1; JOINED.
DR EMBL; AF346448; AAN77251.1; JOINED.
DR EMBL; AF346449; AAN77251.1; JOINED.
DR EMBL; AF346450; AAN77251.1; JOINED.
DR EMBL; AF346451; AAN77251.1; JOINED.
DR EMBL; AF346452; AAN77251.1; JOINED.
DR EMBL; AF346453; AAN77251.1; JOINED.
DR EMBL; AF346454; AAN77251.1; JOINED.
DR EMBL; AF346455; AAN77251.1; JOINED.
DR EMBL; AF346456; AAN77251.1; JOINED.
DR EMBL; AF346457; AAN77251.1; JOINED.
DR EMBL; AF346458; AAN77251.1; JOINED.
DR EMBL; AF346459; AAN77251.1; JOINED.
DR EMBL; AF346460; AAN77251.1; JOINED.
DR EMBL; AF346461; AAN77251.1; JOINED.
DR EMBL; AF346462; AAN77251.1; JOINED.
DR EMBL; AF346463; AAN77251.1; JOINED.
DR EMBL; AF346464; AAN77251.1; JOINED.
DR EMBL; AY431161; AAN38831.1; ALT_SEQ.
DR EMBL; AK50390; BAC34222.1; ..
DR EMBL; AK380024; BAC37808.1; ..
DR XGD; MGI:109151; Lbbp.
DR InterPro; IPR000152; Asx_hydroxy-
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR036209; EGF_Like.
DR InterPro; IPR002212; Fibril-assoc.
DR InterPro; IPR006210; EGF.
DR Pfam; PF00228; EGF; 15.
DR Pfam; PF00683; TS; 4.
DR SMART; SM00181; EGF; 17.
DR SMART; SM00179; EGF_CA; 16.
DR PROSITE; PS00010; ASX_HYDROXY-; 13.
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS01186; EGF_2; 11.
DR PROSITE; PS01187; EGF_CA; 15.
DR Growth factor binding; Repeat; EGF-like domain;
KW Hydroxylation; Signal; Glycoprotein;
KW Alternative splicing.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 1389 LATENT TRANSFORMING GROWTH FACTOR BETA
FT DOMAIN 295 335 BINDING PROTEIN, ISOFORM 1S.
FT REPEAT 343 408 EGF-LIKE 1, CALCIUM-BINDING
FT DOMAIN 542 583 (POTENTIAL).
FT DOMAIN 584 625 EGF-LIKE 2, CALCIUM-BINDING
FT DOMAIN 626 666 EGF-LIKE 3, CALCIUM-BINDING
FT DOMAIN 667 706 EGF-LIKE 4, CALCIUM-BINDING
FT DOMAIN 707 747 EGF-LIKE 5, CALCIUM-BINDING
FT DOMAIN 748 788 EGF-LIKE 6, CALCIUM-BINDING
FT DOMAIN 789 829 EGF-LIKE 7, CALCIUM-BINDING
FT DOMAIN 830 870 EGF-LIKE 8, CALCIUM-BINDING
FT DOMAIN 871 912 EGF-LIKE 9, CALCIUM-BINDING
FT DOMAIN 913 954 EGF-LIKE 10, CALCIUM-BINDING
FT DOMAIN 955 997 EGF-LIKE 11, CALCIUM-BINDING
FT REPEAT 1012 1079 EGF-LIKE 12, CALCIUM-BINDING
FT DOMAIN 1032 1134 REPEAT B.
FT DOMAIN 1135 1175 EGF-LIKE 13, CALCIUM-BINDING
FT REPEAT 1185 1257 EGF-LIKE 14, CALCIUM-BINDING
FT DOMAIN 1289 1330 REPEAT C.
FT DOMAIN 1330 1374 EGF-LIKE 15.
FT DISULFID 299 310 EGF-LIKE 16, CALCIUM-BINDING
FT DISULFID 305 319 (POTENTIAL).
FT DISULFID 321 334 POTENTIAL.
FT DISULFID 346 358 POTENTIAL.
FT DISULFID 553 567 POTENTIAL.
FT DISULFID 569 582 POTENTIAL.
FT DISULFID 588 600 POTENTIAL.
FT DISULFID 595 609 POTENTIAL.
FT DISULFID 611 624 POTENTIAL.
FT DISULFID 630 641 POTENTIAL.
FT DISULFID 636 650 POTENTIAL.
FT DISULFID 653 665 POTENTIAL.
FT DISULFID 671 682 POTENTIAL.
FT DISULFID 677 691 POTENTIAL.

```

FT DISULFID 694 705 POTENTIAL.
FT DISULFID 711 722 POTENTIAL.
FT DISULFID 717 731 POTENTIAL.
FT DISULFID 733 746 POTENTIAL.
FT DISULFID 752 763 POTENTIAL.
FT DISULFID 758 772 POTENTIAL.
FT DISULFID 774 787 POTENTIAL.
FT DISULFID 793 804 POTENTIAL.
FT DISULFID 799 813 POTENTIAL.
FT DISULFID 815 828 POTENTIAL.
FT DISULFID 834 846 POTENTIAL.
FT DISULFID 841 855 POTENTIAL.
FT DISULFID 857 869 POTENTIAL.
FT DISULFID 875 887 POTENTIAL.

Query Match 10.7%; Score 601; DB 1; Length 1389;
Best Local Similarity 24.2%; Pred. No. 2,2e-32;
Matches 230; Conservative 105; Mismatches 305; Indels 310; Gaps 52;

QY 45 DVDECAQGLDCHADALCQNTPTSYKCSCKPGYQG--EGROCEDIDECQNE---LNGGCV 99
Db 584 DIDECAQVRLHC--SQRCENTESFLVCVPAGMASEEGTNCIDVDCLRPDMCRDGRCI 642
QY 100 H-----DCLNIPGNRYCT-CPDGFMLAH 121
Db 643 NTAGAFRCYCDSGYRMRRGYCEDIDELKPTCEEQVNTPGSYQCVPTCEGFR--- 693
QY 122 DGHN--CLEYDECLENNGGCHQTCVNVMSYECCKEGFFLSDNQHTC--IHRSEGLSC 177
Db 700 -GNGQCLDVDECLQPKVCTNGSCTNLEGSYMCSCHRGYSPTPEHRHCCDIDECQGNLC 759
QY 178 MNKDGHGCHICKAPRGSAECRPGFEL--AKVQRDCILTCNHGNGGQCH--SCDPTADG 234
Db 759 MNGQ-----CRNTD-GSFRCTCGGQVLSAAKDQCED:DECEH-HHLCSHQCRNTEGS 810
QY 235 PEGSCRPQYKMTJGRSCLREBTIVLEVTESNTTSVVQSGKRVKRRLLMETCAVNGGCC 294
Db 911 FQVCVQGYRASVLGDHC---ED-----INE-----CLEDSSVCQG-- 844
QY 295 RTCKDSTGVHSCPVGFTQLQDKTKCIDECOTRNGGCHP--CKNTVGSFDCGCKXG 352
Db 845 -DCINTAGSDTCTCPDGFQNL-DNKGQDLINE-AQGLCGSHGECNTQGSFHCVCBEG 901
QY 353 FKLLTDEKSCQDVDECSLDRTCD-HS-CINHPGTACACNRYTL-YGTHCGDNTNCSI 409
Db 902 FSISADGRTCEDIDECVNNVTCDSHGFCNTAGSFRLCVQGFQAPDGGCVQDVNECEL 961
QY 410 NNGGC-QQCVNTVGSYECOC---HPGY-----KLNHNKDC---V 443
Db 962 LSGVCGEAFCEVNEGSGFLVCADENQESYPMTCQCRSVTSGVDVDRQPREKKECYVNL 1021
QY 444 EVKGLLPTSVSRVSLH--CGKSGGGDGCFLRCHSGIHLSSDVTIRTSTVKLNKKGCS 501
Db 1022 NDASLCDNVLAPNVTQCECCTSGAGW----- 1049
QY 502 LKNAELPP-----EGLRPALPKHSVSKESFR-----YNLTCSSCK 538
Db 1050 -DNCEIFPCPVQGTAEFTMCPKGLVPAGESSYDTGGENYKDADECLLFGEEICKNGY 1108
QY 539 QVPGAPCRPTPKEMFTIVFELETNQKVTASCDLSC-VKRTKRLRAIKR-LRKAVER 598
Db 1109 CLINTQPG-----YECYCKQGTYYDPVKLQC----- 1133
QY 599 EQFHLSQGNWLDVAKKPPRTSRQAESCGVGOGGHAENOCVS-----CRAGTYDGA 650
Db 1134 -----FDMDECOQP-----NSC-----IDGQCWNTEGSYNCFCTHPWLDAS 1170
QY 651 RERCILCPNPTFQNESQMTCEBC-----PRPNSGALYT-----PEANWSE 693
Db 1171 EXRCVQPTSENEQTEEDVYVQDLQWEHLSEYVCSRLPVKGQTTYECCLYGEAGWM-Q 1229
QY 694 CGGLCQPGSEYSDGFAFCQLCALGTGFOPEAGRTSCF-----PCGGGLATKHQ 740

```

```

Db 1230 C-ALCPMKD--SDYA--QLCNI---PVTGRRRPYGRALVDFSRQYGPETDPVFICQR 1280
QY 741 GATSFQDCETRVQCSPOHFFYNTTHRCIRCPVG-----TYQFEGKXNCVSCPNTTIT 793
Db 1281 FLNSSFELQAE-EC--GILNGCNGRCVRVHEGYTCCDFDGYHLDMAKTCVDV--NECS 1335
QY 794 DFGSTNITQCNRRCGGELGDF-----TGVIRES--PNY--PGNYPANTE 834
Db 1336 ELNNRMSL--CKNAKCNITGGSYKCLCLPGYIPSDKPNYCTPLNSALNLD 1383

RESULT 12
LTBL_MOUSE
ID LTBL_MOUSE STANDARD: PRT: 1713 AA.
AC Q8CGL9; Q88349; Q8BNW7; Q8C7P5; Q8CIR0;
DT 15-SEP-2003 (Rel. 42, Created:
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Latent transforming growth factor beta binding protein, isoform 1L
DE precursor (LTBP-1) (Transforming growth factor beta-1 binding protein
DE 1) (TGF-beta-1-BP-1).
CN LTBP1.
OS Mus musculus (Mouse).
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Rodentia; Sciurognathia; Muridae; Murinae; Mus
CC NCBI_TaxID=10090;
RN 1;
RP SEQUENCE FROM N.A.
RX PubMed=1271388;
RA Noguera I., Obata H., Gualandris A., Cowie P., Rifkin D.B.;
RT "Molecular cloning of the mouse Ltbp-1 gene reveals a tissue specific
RT expression of alternatively spliced forms.";
R1 Gene 306:31-41 (2003);
R2 [2];
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RX PubMed=1271389;
RA Weiskirchen R., Moser M., Guenther K., Weiskirchen S., Gressner A.M.;
RT "The murine latent transforming growth factor-beta binding protein
RT (Ltbp-1) is alternatively spliced, and maps to a region syntenic to
RT human chromosome 2p21-22.";
RL Gene 308:41-52 (2003);
R3 [3];
RP SEQUENCE OF 1112-1713 FROM N.A.
RC STRAIN=C57BL/6J;
RX TISSUE=Aorta, Liver, and Vein;
RP PubMed=12466951;
RA Okazaki Y., Furuno Y., Kasukawa T., Adachi C., Bono H., Kondo S.;
RA Nikaide I., Otsu N., Sato K., Suzuki H., Yamakawa I., Kiyosawa H.;
RA Yagi K., Toraru Y., Hasegawa Y., Nogami A., Schorbach C., Gotohori T.,
RA Badarrelli R., Hill D.P., Butt C., Hume D.A., Quackenbush J.;
RA Schriell L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.;
RA Blake C.A., Bradt D., Brusci V., Chothia C., Corbani J.E., Cousins S.,
RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.;
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.;
RA Grimmond S., Gustincich S., Hirokawa N., Jackson J.J., Jarvis E.D.,
RA Kanai A., Kawai H., Kawasawa Y., Kedzietski R.M., King B.L.;
RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.;
RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.;
RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.;
RA Petravsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.;
RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring S.Z., Ringwald M.;
RA Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.;
RA Sultana R., Takeraka Y., Taylor M.S., Teasdale R.D., Tomita K.;
RA Verardo J.G., Wynshaw-Boris A., Wang Y., Watanabe Y., Wells C.;
RA Wilming J.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.;
RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayashi N.;
RA Hirozane-Kisikawa I., Konno H., Nakamura M., Sakazume N., Sato K.;
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.;
RA Hara A., Hashizume K., Iotani K., Ishii Y., Itoh M., Kagawa T.;
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.;
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.;
RA Birney E., Hayashizaki Y.;

```



```

Db 908 DIDECAQVRLHC-SQRCENTGSLFVCPAGFMASEEGTNCIDVDECLRPDMCDGRGI 966
QY 100 H-----DCLNIPGNRYCT-CFDFMLAH 121
Db 967 NTAGAFRCYCDSGYRMSRGYCEDIDELKFTCPREEQCVNTPSQYCVPTCEGFR--- 1023
QY 122 DGHN--CLDVBDECLNNGGQHTCVNMVMSYSCCKKEGFFLSDNOHTC--IHRSEGLSC 177
Db 1024 -GNGGQCLDVBDECLKPVCTNGSCTNLEGSYMCSTRGYSPTPDHRHCQIDECQCGNLC 1082
QY 178 MNKHGCHSHICKEAPRGVACRCPGEL--AKNORDCILTCNKGNGCQH--SCDSTADG 234
Db 1083 MNGQ-----CRNTD-GSFRCTCGGYQLSAKQCEDIDECHE-HHLCSHGQCRNTSGS 1134
QY 235 PEGSCHPYQMYHTDGRSCLEREDVLEVTESNTTSVVDGKVRKRRJLMTCAVNNGCCD 294
Db 1135 FQVCVNGYRASVLGDHC---ED-----INE-----CLEDSVCQGG-- 1168
QY 295 RTCKDSTGVHSCRPVGFLLDQKTKDIDECOTRNGGCDHF--CKNIVUSFDCGCKKG 352
Db 1169 -DCINTAGSYDCTCPGQFLN-DNKGQDINBC-AQPLGCGSHGECNLQGSFHCVCQG 1225
QY 353 FKLLTDEKSCQVDECSLDRDTC-HS-C-NHRTGTACACNRGYTL-VGPTHGCDTNECS 409
Db 1225 FS-SAGRTCEDIDECVNTVCDSHGFCNTAGSPRLCYQGFQAPQDQCGCVVNECEL 1285
QY 410 NNGGC-QQVCVNTVGYEQC---HPGY-----KLNHKKQCG---V 443
Db 1286 LSGVCGEAFCEVNEGFLCVACDENGEYSPMTGQCSRVTEDSGVDROPREEKCEYCNL 1345
QY 444 EVKGLLPTSVSRVSLH--CGKSGGDDGCLFCHSGIHLSSDVTIRTSTVFKLNEGKCS 501
Db 1346 NDASJCDNVLPVNTVQECCTSGAGWG----- 1373
QY 502 LKNAELFP-----EGLRALPKHSVKSFSR-----VYNLTCSGK 538
Db 1374 -DNCEIFPCPVQGTAEFTBMCPRGKGLVPAGESYDTGGENYKQADCELLFGEELCKNGY 1432
QY 539 QVPSAGRSTPKEMETIVFELETNKEVTVASCDLSCLVKTETKRLKRLTKLKAHVR 598
Db 1433 CLNTQSG-----YECYKQGYYYPVVKLQ----- 1457
QY 599 EQPHLQSGMNLVDVAKPKPTSRERQAESCGVGQGHAEQCVS-----CRAGTYDGA 650
Db 1458 -----FQWDECCDP-----NSC-----IDGQCVNTEGSCYNCFCTHPVULAS 1494
QY 651 RERCILCPNGTFQNEEGQWTCPC-----PRPNSGALKT-----PERANNSE 693
Db 1495 ERKCVQFTESNEQIEETDYYQLCNEHLSEYVCSRPLVKGCTTYTECCCLYGEAWGM-Q 1553
QY 694 CGGLQCPGYSADGAPCALGTQFPBAGTSCF-----PCGGGLATKHQ 740
Db 1554 A-ALCPMKD--SDYA--QLCNI-----PVTGRRPYGRDALVDFSEQVGPETDPYFQDR 1604
QY 741 GATSPQDCETRVQCSQGHFYNTTHRCIRCPVG-----TYQEPFGKNNVCSEGNNTT 793
Db 1605 FLMSFEELQAE-EC--GILGNGCNGRCVRVHEGYTCDFGQVHLDMAKMTCDV--NECS 1659
QY 794 DFDGNTNITCKNRRCGGELGDF-----TGYES--PNY--PCNYPANTE 834
Db 1660 ELNRMSL--CKNAKINTGYSYKCLCPGLPYSPDKPNYCTPLNSALNLD 1707

RESULT 13
LTBP-1
ID LTBP-1 RAT STANDARD; PRAT: 1712 AA.
AC Q00918;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Latent transforming growth factor beta binding protein 2 precursor
DE (LTBP-1) (Transforming growth factor beta-1 binding protein: 1) (TGF-

```

```

DE betal-BP-1) (Transforming growth factor beta-1 masking protein, large
DE subunit).
GN LTSP1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [-]
SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=91062373; PubMed=2247454;
RA Tsuji T., Okada F., Yamaguchi K., Nakamura T.
RT "Molecular cloning of the large subunit of transforming growth factor
RT type beta masking protein and expression of the mRNA in various rat
RT tissues."
RL Proc. Natl. Acad. Sci. U.S.A. 87:8835-8839(1990).
CC -1- SUBUNIT: THE LARGE LATENT COMPLEX OF TGF-BETAL FROM PLATELETS IS
CC COMPOSED OF THE TGF-BETAL MOLECULE NONCOVALENTLY ASSOCIATED WITH
CC A MASKING PROTEIN CONSISTING OF A DISULFIDE-BONDED COMPLEX OF A
CC DIMER OF THE N-TERMINAL PROPEPTIDE OF THE TGF-BETAL PRECURSOR AND
CC A THIRD COMPONENT DENOTED TGF-BETAL-BP (OR MP LARGE SUBUNIT).
CC TGF-BETAL-BP DOES NOT BIND DIRECTLY TO ACTIVE TGF-BETAL.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: Contains 18 EGF-like domains.
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation.
CC The European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EVBL; M5543; AAA-2235.1; --
CC PIR: A38261; A38261.
CC HSSP; P:6139; 1FSB.
CC InterPro; IPR003152; Asx hydroxyl.
CC InterPro; IPR001881; EGF-Ca.
CC InterPro; IPR006259; EGF-like.
CC InterPro; IPR002212; Fibfil-assoc.
CC Pfam; PF00005; EGF; 16.
CC Pfam; PF00693; TB; 4.
CC SMART; SM00179; EGF_CA; 13.
CC ProSITE; PS00010; ASX HYDROXYL; 13.
CC ProSITE; PS00022; EGF_1; 2.
CC ProSITE; PS01186; EGF_2; 10.
CC ProSITE; PS01187; EGF_CA; 15.
KW Growth factor binding; Repeat; EGF-like domain; Hydroxylation; Signal.
FT SIGNAL 1 20
FT PROPEP 21 736
FT CHAIN 737 1577
FT
FT S-TE 734 736
FT SITE 1575 1577 CLEAVAGE (POTENTIAL).
FT PROPEP 1578 1712 CLEAVAGE (POTENTIAL).
FT DOYAIN 181 213 POTENTIAL.
FT DOYAIN 391 423 EGF-LIKE 1.
FT REPEAT 551 604 EGF-LIKE 2.
FT DOYAIN 618 658 INTERNAL REPEAT 1.
FT REPEAT 671 721 INTERNAL REPEAT 2.
FT DOYAIN 865 966 EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).
FT DOYAIN 907 949 EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).
FT DOYAIN 949 989 EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT DOYAIN 990 1029 EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).
FT DOYAIN 1030 1070 EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT DOYAIN 1071 1111 EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT DOYAIN 1112 1152 EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT DOYAIN 1153 1193 EGF-LIKE 10, CALCIUM-BINDING (POTENTIAL).
FT DOYAIN 1194 1235 EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT DOYAIN 1236 1277 EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT DOYAIN 1278 1320 EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT REPEAT 1340 1392 EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT DOYAIN 1415 1457 INTERNAL REPEAT 3.
FT DOYAIN 1458 1498 EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT DOYAIN 1459 1498 EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).

```

FT	REPEAT	1517	1568	INTERNAL REPEAT 4.
FT	DOMAIN	1612	1652	EGF-LIKE 17.
FT	DOMAIN	1653	1697	EGF-LIKE 18. CALCIUM-BINDING (POTENTIAL).
FT	D:SULFID	185	195	BY SIMILARITY.
FT	D:SULFID	189	201	BY SIMILARITY.
FT	D:SULFID	203	212	BY SIMILARITY.
FT	D:SULFID	395	405	BY SIMILARITY.
FT	D:SULFID	399	411	BY SIMILARITY.
FT	D:SULFID	413	422	BY SIMILARITY.
FT	D:SULFID	622	633	BY SIMILARITY.
FT	D:SULFID	628	642	BY SIMILARITY.
FT	D:SULFID	644	657	BY SIMILARITY.
FT	D:SULFID	869	881	BY SIMILARITY.
FT	D:SULFID	876	890	BY SIMILARITY.
FT	D:SULFID	892	905	BY SIMILARITY.
FT	D:SULFID	911	923	BY SIMILARITY.
FT	D:SULFID	918	932	BY SIMILARITY.
FT	D:SULFID	934	947	BY SIMILARITY.
FT	D:SULFID	953	964	BY SIMILARITY.
FT	D:SULFID	959	973	BY SIMILARITY.
FT	D:SULFID	976	988	BY SIMILARITY.
FT	D:SULFID	994	1005	BY SIMILARITY.
FT	D:SULFID	1000	1014	BY SIMILARITY.
FT	D:SULFID	1017	1028	BY SIMILARITY.
FT	D:SULFID	1034	1045	BY SIMILARITY.
FT	D:SULFID	1040	1054	BY SIMILARITY.
FT	D:SULFID	1056	1069	BY SIMILARITY.
FT	D:SULFID	1075	1086	BY SIMILARITY.
FT	D:SULFID	1081	1095	BY SIMILARITY.
FT	D:SULFID	1097	1110	BY SIMILARITY.
FT	D:SULFID	1116	1127	BY SIMILARITY.
FT	D:SULFID	1122	1136	BY SIMILARITY.
FT	D:SULFID	1138	1151	BY SIMILARITY.
FT	D:SULFID	1157	1169	BY SIMILARITY.
FT	D:SULFID	1164	1178	BY SIMILARITY.
FT	D:SULFID	1180	1192	BY SIMILARITY.
FT	D:SULFID	1198	1210	BY SIMILARITY.
FT	D:SULFID	1204	1219	BY SIMILARITY.
FT	D:SULFID	1221	1234	BY SIMILARITY.
FT	D:SULFID	1240	1252	BY SIMILARITY.
FT	D:SULFID	1246	1261	BY SIMILARITY.
FT	D:SULFID	1263	1276	BY SIMILARITY.
FT	D:SULFID	1282	1294	BY SIMILARITY.
FT	D:SULFID	1289	1303	BY SIMILARITY.
FT	D:SULFID	1305	1319	BY SIMILARITY.
FT	D:SULFID	1419	1432	BY SIMILARITY.
FT	D:SULFID	1427	1441	BY SIMILARITY.
FT	D:SULFID	1443	1456	BY SIMILARITY.
FT	D:SULFID	1462	1473	BY SIMILARITY.
FT	D:SULFID	1468	1482	BY SIMILARITY.
FT	D:SULFID	1484	1497	BY SIMILARITY.
FT	D:SULFID	1616	1627	BY SIMILARITY.
FT	D:SULFID	1622	1636	BY SIMILARITY.
FT	D:SULFID	1638	1651	BY SIMILARITY.
FT	D:SULFID	1657	1672	BY SIMILARITY.
FT	D:SULFID	1667	1681	BY SIMILARITY.
FT	D:SULFID	1693	1696	BY SIMILARITY.
FT	CARBOHYD	339	370	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	370	416	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	416	416	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	612	612	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1042	1042	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1242	1242	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1357	1357	N-LINKED (GLCNAC. .) (POTENTIAL).
SQ	SSEQUENCE	1712 AA;	186598 MW; 650BCEAA691FD134 CRC64;	

Query Match 10.6%; Score 592; DB 1; Length 1712;
Best Local Similarity 24.3%; Pred. No. 1,le-31;
Matches 230; Conservative 102; Mismatches 305; Indels 310; Gaps 52;

QY	45	DVDECAQGLDDHADALCONTPTSYKSCCKPGY--QGEGRCQCEIDDEC-----GNEI 94
DB	907	DIDCAQAQHLCC-SQRCENTEGSFUCIPAGFIASEGNSCIDVDECLRPVDCRGRCI 965

QY	95	N-----GGC-----VHDCINIPGNRYCT-CFDCGFMJAH 121
DB	966	NTAGAFRCYCDGYSRMSRRGHGCE-DECLTPGTCPEEQVNSPGSYCCVPTCEGER--- 1022
QY	122	DGHN--CLDYDECLNNGCGOHTCVNMVMSYECCKCKEFLSDNQHCTC--IHRSEGLSC 177
DB	1023	-GNGQCJEDVDECLQPKVCTNGSCTNLEGSYKSCCHKGYSPTPCHRHCDIDECQGNJC 1081
QY	178	VXKDHGCHICKAPRGSVACEKCPGFEI--AKNQRDCILTCNWHGNGGCGH--SCDIDAOG 234
DB	1082	MNGO-----CKNTD--GSPFCTCGQYQLSAKDQCEIDCECHRH--LSHGCRNTEBS 1133
QY	235	PEGCHPQYKCHTQGRSCLEREDTVLEVTESENTTSVVGDGRVKRRLJLMEETCAVNGGCD 294
DB	1134	FQCLNQGYAASVLGDHC---ED---INE-----CLEDSSVCQGG-- 1167
QY	295	RTCKDTSTGVHSCSPVGFLLQDGKTKDIDECOTRNGGCDHF--CKNVGSFDCCKKG 352
DB	1168	-DCINTAGSYDCTCPDGLQIN-DNKOCODINEC-AQPLGAPHGECINTQGSFHCYCEG 1224
QY	353	FKULTDEKSCQVDECSJDRTCO-HS-CINHPGTFCACNRGVTJ--YGFTHCCTVECSI 409
DB	1225	FSISAGRTCEDIDECUNNTVCDSHGFCNTAGSFCLCYGFCAPDQGGCVDVNECEL 1284
QY	410	NNGGC-QQVCNVTGSGYCOCHPGYKLH-----WNKDC---V 443
DB	1285	LSGVGGEAFCEVNEGSGFLVCADENQESMTQCRSRATEDSGVDRQPKKEKCYNL 1344
QY	444	EVKGLFTSVSPRVSJH--CGKSGGDDGFLRCHSGIHLSSTVTTTTSVTFKLNEGKS 591
DB	1345	NDASLDCNVLPNVTQBOCCTSGAGW-----EGELRALPEKHSVSKESFR----- 1372
QY	502	LKNAELFP-----EGELRALPEKHSVSKESFR-----VNLTCSSGK 538
DB	1373	-DNCEIFPCPVGTABEFSEMCPRGKGFVPAGESSEYETGGENYKADABECULFGEI 1431
QY	539	QVPGAPGRPTPKEMFITVEFELETNOKEVTASCDLSQIVKTEKRLKXALRLKXAVHR 598
DB	1432	C.LNTQGYECYCKE-----GTYYDPVKLQC----- 1456
QY	599	EQPHQLSGNLDVAKKPARTSERQAESGVGGHAENQCVS-----CRAGTYDGA 650
DB	1457	-----FMDDECQDP-----NSC-----IDGCVNTEGSYNCFCTHPVYLDAS 1493
QY	651	RERCILCPNGTFOKEGQMTCEPC-----PRGNSGALKT-----PEANVMSR 693
DB	1494	EKRCVQPTESNBOIEETDVYQDLQWEHLSEYVCSRPJVGKQITYTECCCLYGEAWGM-Q 1552
QY	694	CGGLCQGEYSADGAPACQJCALGTTQPEAGRTSCF-----PCGGGLATKHQ 740
DB	1553	C-ALCEMKD--SDQYA--QLCNI-----PVTGRRRPYGRDALVDFSEQYGPETDPYFIQDR 1603
QY	741	GATSFQDCETRVQCSGCHFYNTTHRCICRPVG-----TYQPEGKNKNCUSCPNTTIT 793
DB	1604	FLANSFEELQAE-EC--GILNGCENGRCVRQESYTCDFDGHLDMAKTCVDV--NECS 1658
QY	794	DFDGSNITCKNRKCGGELGDF-----GYIES--PNVPGNYNPANT 832
DB	1659	EJNNRSL--CKNAKCINTEGSKYKCVCLFGYVPSDKPNY--CTPLNT 1701

RESULT 14

FBL2 MOUSE

ID FBL2 MOUSE STANDARD; PRF: 1221 AA.

AC P37855; Q3WU12;

DT 01-OCT-1994 (Rel. 36, Created)

DT 01-OCT-1994 (Rel. 30, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE Fblulin-2 precursor.

SN FBLN2.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 27-35.
 RC TISSUE=Fibroblast;
 RX MEDLINE=94064787; PubMed=62451130;
 RA Pan T.-C., Sasaki T., Zhang R.-Z., Faessler R., Timpl R., Chu M.-J.;
 "Structure and expression of fibulin-2, a novel extracellular matrix
 protein with multiple EGF-like repeats and consensus motifs for
 calcium binding.";
 RT J. Cell Biol. 123:1269-1277(1993).
 RN [2]
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
 RX MEDLINE=99337686; PubMed=10406956;
 RA Graessel S., Sicot F.-X., Gotta S., Chu M.-L.;
 "Mouse fibulin-2 gene. Complete exon-intron organization and promoter
 characterization.";
 RT Eur. J. Biochem. 263:471-477(1999).
 RN [3]
 RP DEVELOPMENTAL STAGE.
 RX PubMed=850569;
 RA Zhang H.-Y., Timpl R., Sasaki T., Chu M.-L., Ekblom P.;
 "Fibulin-1 and fibulin-2 expression during organogenesis in the
 developing mouse embryo.";
 RT Dev. Dyn. 205:348-364(1996).
 RN [4]
 RP BINDING TO LAMA2.
 RX PubMed=1002829;
 RA Talts J.F., Andac Z., Goshring W., Brancaccio A., Timpl R.;
 "Binding of the G domains of laminin alpha1 and alpha2 chains and
 perlecan to heparin, sulfatides, alpha-dystroglycan and several
 extracellular matrix proteins.";
 RT EMBO J. 18:863-870(1999).
 RN [5]
 RP DOWN-REGULATION BY GLUCOCORTICOID.
 RX PubMed=11737251;
 RA Gu Y.-C., Talts J.F., Gullberg D., Timpl R., Ekblom M.;
 "Glucocorticoids down-regulate the extracellular matrix proteins
 fibronectin, fibulin-1 and fibulin-2 in bone marrow stroma.";
 RT Eur. J. Haematol. 67:176-184(2001).
 CC -!- FUNCTION: ITS BINDING TO FIBRONECTIN AND SOME OTHER LIGANDS IS
 CALCIUM DEPENDENT.
 CC -!- SUBUNIT: Homotrimer; disulfide-linked. Interacts with LAMA2.
 CC -!- SUBCELLULAR LOCATION: Secreted; extracellular matrix.
 CC -!- ALTERNATIVE PRODUCTS:
 Event=Alternative splicing; Named isoforms=2;
 Comment=Additional isoforms seem to exist;
 Name=1;
 IsoId=P37889-1; Sequence=Displayed;
 Name=2; Synonyms=EGF3-less;
 IsoId=P37889-2; Sequence=VSP 001391;
 CC -!- TISSUE SPECIFICITY: Component of both basement membranes and other
 connective tissues.
 CC -!- DEVELOPMENTAL STAGE: The differential expression of the fibulin
 family contributes to the formation of molecularly distinct
 extracellular matrices already during early developmental stages
 of a large number of tissues.
 CC -!- INDUCTION: Glucocorticoids suppressed mRNA expression and protein
 synthesis.
 CC -!- SIMILARITY: Belongs to the fibulin family.
 CC -!- SIMILARITY: Contains 3 anaphylatoxin-like domains.
 CC -!- SIMILARITY: Contains 11 EGF-like domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X75285; CAAS3040.1;
 DR EMBL; AF135253; AAD34456.1; -;
 DR EMBL; AF135253; AAD34456.1; -;

DR EMBL; AF135239; AAD34456.1; JOINED.
 DR EMBL; AF135240; AAD34456.1; JOINED.
 DR EMBL; AF135241; AAD34456.1; JOINED.
 DR EMBL; AF135242; AAD34456.1; JOINED.
 DR EMBL; AF135243; AAD34456.1; JOINED.
 DR EMBL; AF135244; AAD34456.1; JOINED.
 DR EMBL; AF135245; AAD34456.1; JOINED.
 DR EMBL; AF135246; AAD34456.1; JOINED.
 DR EMBL; AF135247; AAD34456.1; JOINED.
 DR EMBL; AF135248; AAD34456.1; JOINED.
 DR EMBL; AF135249; AAD34456.1; JOINED.
 DR EMBL; AF135250; AAD34456.1; JOINED.
 DR EMBL; AF135251; AAD34456.1; JOINED.
 DR EMBL; AF135252; AAD34456.1; JOINED.
 DR FIR; A49457; A49457.
 DR HSP; P0736; IAPQ.
 DR MGI; 95488; Pbln2.
 DR InterPro: IPR000020; Anaphylatoxin.
 DR InterPro: IPR000152; Asx hydroxyl.
 DR InterPro: IPR001881; EGF_Ca.
 DR InterPro: IPR006209; EGF_Like.
 DR Pfam: PF01821; ANATO; 2.
 DR Pfam: PF00308; EGF; 5.
 DR SMART: SM00104; ANATO; 3.
 DR SMART: SM00179; EGF_Ca; 9.
 DR PROSITE: PS00010; ASX HYDROXYL; 5.
 DR PROSITE: PS01177; ANAPHYLATOXIN_1; 3.
 DR PROSITE: PS01178; ANAPHYLATOXIN_2; 3.
 DR PROSITE: PS00022; EGF_1; FALSE_NEG.
 DR PROSITE: PS01196; EGF_2; 5.
 DR PROSITE: PS01187; EGF_Ca; 10.
 KW Signal; Glycoprotein; Extracellular matrix; Plasma; EGF-like domain;
 KW Calcium-binding; Alternative splicing; Repeat.
 FT SIGNAL 1 26
 FT CHAIN 27 1221 FIBULIN-2.
 FT DOMAIN 27 434 N.
 FT DOMAIN 27 434 SUBDOMAIN NA (CYS-RICH).
 FT DOMAIN 177 434 SUBDOMAIN NB (CYS-FREE).
 FT DOMAIN 435 477 ANAPHYLATOXIN-LIKE 1.
 FT DOMAIN 478 510 ANAPHYLATOXIN-LIKE 2.
 FT DOMAIN 511 543 ANAPHYLATOXIN-LIKE 3.
 FT DOMAIN 594 635 EGF-LIKE 1, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 669 708 EGF-LIKE 2.
 FT DOMAIN 709 755 EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 756 800 EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 801 846 EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 847 894 EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 895 937 EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 938 979 EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 980 1018 EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 1019 1061 EGF-LIKE 10, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 1062 1106 EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 1111 1221 DOMAIN III.
 FT SITE 421 423 CELL ATTACHMENT SITE (POTENTIAL).
 FT DISULFID 435 462 BY SIMILARITY.
 FT DISULFID 436 469 BY SIMILARITY.
 FT DISULFID 449 470 BY SIMILARITY.
 FT DISULFID 479 508 BY SIMILARITY.
 FT DISULFID 492 509 BY SIMILARITY.
 FT DISULFID 511 535 BY SIMILARITY.
 FT DISULFID 512 543 BY SIMILARITY.
 FT DISULFID 525 543 BY SIMILARITY.
 FT DISULFID 598 610 BY SIMILARITY.
 FT DISULFID 606 619 BY SIMILARITY.
 FT DISULFID 621 634 BY SIMILARITY.
 FT DISULFID 673 683 BY SIMILARITY.
 FT DISULFID 673 692 BY SIMILARITY.
 FT DISULFID 694 707 BY SIMILARITY.
 FT DISULFID 713 726 BY SIMILARITY.
 FT DISULFID 723 735 BY SIMILARITY.
 FT DISULFID 742 754 BY SIMILARITY.
 FT DISULFID 805 818 BY SIMILARITY.
 FT DISULFID 812 827 BY SIMILARITY.


```
FT DISULFID 833 845 BY SIMILARITY.
FT DISULFID 899 912 BY SIMILARITY.
FT DISULFID 906 921 BY SIMILARITY.
FT DISULFID 923 936 BY SIMILARITY.
FT DISULFID 942 954 BY SIMILARITY.
FT DISULFID 950 963 BY SIMILARITY.
FT DISULFID 965 978 BY SIMILARITY.
FT DISULFID 984 993 BY SIMILARITY.
FT DISULFID 989 1002 BY SIMILARITY.
FT DISULFID 1004 1017 BY SIMILARITY.
FT DISULFID 1023 1035 BY SIMILARITY.
FT DISULFID 1031 1044 BY SIMILARITY.
FT DISULFID 1046 1060 BY SIMILARITY.
FT DISULFID 1066 1079 BY SIMILARITY.
FT DISULFID 1073 1088 BY SIMILARITY.
FT DISULFID 1093 1105 BY SIMILARITY.
FT CARBOHYD 179 179 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 497 497 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 737 737 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1072 1072 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPIC 709 755 Missing (in isoform 2).
FT FTID-VSP 001391.
FT CONFLICT 140 159 HSRKKAAGTHTVHSSCRAC -> TVAVSICWFPYAPLILP
FT FT CONFLICT 348 348 S -> L (IN REF. 2).
FT CONFLICT 507 507 Q -> QO (IN REF. 2).
FT CONFLICT 1102 1102 Q -> E (IN REF. 2).
SQ SEQUENCE 1221 AA; 131818 MW; 87DB2A10A8FDC45F CRC64;

Query Match
Best Local Similarity 29.0%; Pred. No. 7.1e-31;
Matches 157; Conservative 66; Mismatches 182; Indels 137; Gaps 29;

QY 44 EDVDECAQGLDDCHADALCONTPTSK-----SKPGY-QEGGROCEDIDCGNELNGG 97
DB 708 EDQDELMGTHDCSKQKQFCVNTLGSFYCVNHTVLCAGYILNAHRKCVDINECVTLH-T 766

QY 98 C--VHDLNIPGNRYC-----TCDFGMLAHGHNCCDVDECLNNGGCG--HTCVNVMGS 149
DB 767 CTRAHCNVNTPSGFQYKALTCBPYVLT-DG-ECTDVECVT-GTHNCOAGSCQNTKGS 824

QY 150 YEC-----CKEGFPLDNGHTCHRS-----EGLSCMNKHGCHSHCKEAPRGSV 196
DB 825 FYCAQRQRCMDG-FLODPEGNCDINECTSLPEPCRSFGSCINTVG--SYTCQNP---- 877

QY 197 ACRCRPGFELAKNORC--ILTCNHNNGGCOHS--CDDTAGPECCSCHPCYXHTDGRSC 252
DB 878 -LYCGRGYHANESECEVDVNECTGVHRCGEQLCYNLPGSYRCCKFGFORDAFGRTC 936

QY 253 LEREDTVLEVTESNTTSWVDGDKRVKRLMETCAVYNNGG-CORTCKDTSTGVHSCPYG 311
DB 937 ID-----VNECWSPGRLCQHTCENTPGSVYRCSAAG 968

QY 312 FTQLDGKTKDIDECOTRNGGCHFCNKLIVGDFCGCKKGFALLTDEKSCQDVIECSLD 371
DB 969 FLAADGKHGDEVNECTRP--CSQECANIYGYCYCRQYGLAEDGHTCTDIDCAQ3 1026

QY 372 R--TCDSHCNHPGTFACAC-NRGYTLV-GFTHCGDINECSINNGCQ--QVCNNTVGSY 425
DB 1027 AGILCTFRCVNPGSVQACPEGYTNMANGRSCKLDDECALGTHNCSAECHN:CGSF 1086

QY 426 EC---QCHPOY----KLHMKNDKCEV-----KGLJ----- 449
DB 1087 RCLRFDCPPNVRVQTKCERTTQDITECQTSFARITHYQLNFQGLLVPAHIFRIGFA 1146

QY 450 PTVSPRVLHCKSGGGGDCFLRCHSGIHLSSDVTIRTSVTFKLN--EGKCSLKNABL 507
DB 1147 PAFAGDTISITIK-GNEEGYF-----VTRRLNAYTGVVSLQSRVSL 1186

QY 508 FP 509
DB 1187 EP 1188
```

RESULT 15

```
FE11_HUMAN
ID FE11_HUMAN STANDARD; PRT; 703 AA.
AC P23142; P23143; P23144; P37888; Q8TBH8; Q8TBQ5; C9U3R4; C9U4H1;
DT C1-NOV-1991 (Rel. 23, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Fibulin-1 precursor.
GN FBUN1.
CS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN 1;
RP SEQUENCE FROM N.A. (ISOFORMS A; B AND C).
RX MEDLINE=9103426; PubMed=2269669;
RA Argaves W.S., Tran H., Burgess W.H., Dickerson K.;
RT "Fibulin is an extracellular matrix and plasma glycoprotein with
RT repeated domain structure.";
RL J. Cell Biol. 111:3155-3164(1990).
RN 2;
RP SEQUENCE FROM N.A. (ISOFORM D), TISSUE SPECIFICITY, AND INTERACTION
RP WITH FN1 AND FGB.
RX PubMed=9106159;
RA Tran H., Mattei M., Godyna S., Argaves W.S.;
RT "Human fibulin-ID: molecular cloning, expression and similarity with
RT SL-5 protein, a new member of the fibulin gene family.";
RL Matrix Biol. 15:479-493(1997).
RN 3;
RP SEQUENCE FROM N.A. (ISOFORM D).
RX MEDLINE=99253993; PubMed=10318851;
RA Krichevsky A.M., Metzger E., Rosen H.;
RT "Translational control of specific genes during differentiation of
RT HL-60 cells.";
RL J. Biol. Chem. 274:14295-14305(1999).
RN 4;
RP SEQUENCE FROM N.A. (ISOFORM C).
RA Gu J.R., Wan D.F., Zhao X.T., Zhou X.M., Jiang H.Q., Zhang P.P.,
RA Qin W.X., Huang Y., Qiu X.K., Qian L.F., He L.F., Li H.N., Yu Y.,
RA Yu J., Han L.H.;
RT "Novel Human cDNA clones with function of inhibiting cancer cell
RT growth.";
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RN 5;
RP SEQUENCE FROM N.A.
RX MEDLINE=20057165; PubMed=10591208;
RA Dunham T., Hunt A.R., Collins C.E., Bruskewich R., Beare D.M.,
RA Clapp M., Smink J.J., Ainscough R., Almeida J.P., Babbage A.K.,
RA Baguley C., Bailey J., Barlow K.F., Bates K.N., Beasley O.P.,
RA Bird C.P., Blakey S.E., Bridgman A.M., Buck D., Burgess J.,
RA Burrill W.D., Burton J., Carder C., Carter N.P., Chen Y., Clark S.,
RA Clegg S.N., Cobley V.E., Cole C.G., Collier R.E., Connor R.,
RA Corcoran D., Corby K.R., Coville G.J., Cox A.V., Davis J., Dawson E.,
RA Dhami P.D., Dockree C., Dodsworth S.J., Durbin R.M., Ellington A.G.,
RA Evans K.L., Fey J.M., Fleming K., French J., Garner A.A.,
RA Gilbert J.G.R., Goward M.E., Graham D.V., Griffiths M.N.D., Hall C.,
RA Hall R.E., Hall-Tamlyn G., Heathcote R.W., Ho S., Holmes S.,
RA Hunt S.E., Jones M.C., Kershaw C., Kimberley A.M., King A.,
RA Jald G.K., Langford C.F., Leversha X.A., Lloyd C., Lloyd D.M.,
RA Kertay I.D., Mashreghi-Mohammadi M., Matthews L.H., Mccann C.T.,
RA Mcclay J., McLaren S., McMuray A.A., Milne S.A., Mortimore B.J.,
RA Odell C.N., Pavitt R., Pearce A.V., Pearson D., Phillimore B.J.C.,
RA Phillips S.H., Plumb R.W., Ramsay H., Ramsey Y., Rogers J., Ross M.T.,
RA Scott C.E., Sehra H.K., Skuce C.D., Smalley S., Smith M.J.,
RA Soderlund C., Spragon L., Steward C.A., Surston J.E., Swann R.M.,
RA Vaudin M., Wall M., Wallis C.M., Whiteley M.N., Willey D.L.,
RA Williams L., Williams S.A., Williamson H., Wilmer T.E., Wilming J.,
RA Wright C.L., Hubbard T., Bentley D.R., Beck S., Rogers J., Shimizu N.,
RA Minoshima S., Kawasaki K., Sasaki T., Asakawa S., Kudoh J.,
RA Shintani A., Shibuya K., Yoshizaki Y., Aoki K., Mitsuyama S.,
RA Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S., Do A., Do T.,
RA Doran A., Farg F., Fu Y., Hu P., Hua A., Kenton S., La H., Lac H.I.,
```

RA Lewis J., Lewis S., Lin S.-P., Loh P., Malaj E., Nguyen T., Par H.,
 RA Phan S., Qi S., Qian Y., Ray L., Ren C., Shaull S., Sloan D., Song L.,
 RA Wang Q., Wang Y., Wang Z., White J., Willingham D., Wu H., Yao Z.,
 RA Zhan M., Zhang G., Chissoe S., Murray J., Miller N., Minx P.,
 RA Fulton R., Johnson D., Benis G., Bentley D., Bradshaw H., Bourne S.,
 RA Cordes M., Du Z., Fulton L., Goela D., Graves T., Hawkins J.,
 RA Hinds K., Kemp K., Latreille P., Layman D., Ozersky P., Rohlfing J.,
 RA Scheet P., Walker C., Wamsley A., Wohldmann P., Pepin K., Nelson J.,
 RA Korf I., Bedell J.A., Hillier L., Mardis E., Waterston R., Wilson R.,
 RA Emanuel B.S., Shaikh T., Kurahashi H., Saitta S., Budarf M.L.,
 RA McDermid H.E., Johnson A., Wong A.C., Morrow B.E., Edelmann L.,
 RA Kim U.J., Shizuya H., Simon M.I., Dumanski J.P., Peyrard M., Keira C.,
 RA Seroussi E., Fransoni I., Tapia I., Bruder C.E., O'Brien K.P.,
 RA Wilkinson P., Bodenteich A., Hartman K., Hu X., Khan A.S., Jane L.,
 RA Tiliakou Y., Wright H.,
 RT "The DNA sequence of human chromosome 22";
 RL Nature 402:489-495(1999).
 [6]
 RN SEQUENCE FROM N.A. (ISOFORM C).
 RP TISSUE=Brain;
 RC MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Sherman C.M., Schuler G.D.,
 RA Altschul S.F., Zeng B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins S.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Donald M.F., Casavant T.J., Scheetz T.E.,
 RA Brownstein M.J., Udell J.B., Toshiyuki S., Carninci P., Frange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullat S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton S., Ketterman M., Yadan A., Rodriguez S., Sanchez A.,
 RA Whiting X., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Jackson X.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S., Krzywinski M.I., Skalska U., Smal M.A.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 [7]
 RN SEQUENCE OF 1-26 FROM N.A.
 RP Castoldi M., Chu M.-L.;
 RA "Structural and functional characterization of the human and mouse
 RT fibulin-1 gene promoters: role of Sp1 and Sp3";
 RL Biochem. J. 362:41-50(2002).
 [8]
 RN SEQUENCE OF 30-44.
 RP MEDLINE=89354537; PubMed=2527614;
 RA Argaves W.S., Dickerson K., Burgess W.H., Ruoslahti E.;
 RT "Fibulin, a novel protein that interacts with the fibronectin
 RT receptor beta subunit cytoplasmic domain";
 RL Cell 58:623-629(1989).
 [9]
 RN SELF-ASSOCIATION AND INTERACTION WITH FN1.
 RP PubMed=1400330;
 RA Balbona K., Tran H., Godyna S., Ingham K.C., Strickland D.K.,
 RA Argaves W.S.;
 RT "Fibulin binds to itself and to the carboxyl-terminal heparin-binding
 RT region of fibronectin";
 RL J. Biol. Chem. 267:20120-20125(1992).
 [10]
 RN POSSIBLE FUNCTION.
 RP PubMed=7534784;
 RA Roark E.F., Keene D.R., Haudenschild C.C., Godyna S., Little C.D.,
 RA Argaves W.S.;
 RT "The association of human fibulin-1 with elastic fibers: an
 RT immunohistological, ultrastructural, and RNA study";
 RL J. Histochem. Cytochem. 43:401-411(1995).
 [11]
 RN INTERACTION WITH FGB.
 RP

RX PubMed=7642629;
 RA Tran H., Tanaka A., Litvinovich S.V., Meved L.V., Haudenschild C.C.,
 RA Argaves W.S.;
 RT "The interaction of fibulin-1 with fibrinogen. A potential role in
 RT hemostasis and thrombosis";
 RL J. Biol. Chem. 270:19459-19464(1995).
 [12]
 RN DEVELOPMENTAL STAGE.
 RP PubMed=8737292;
 RA Miosge N., Gotz W., Sasaki T., Chu M.-L., Timpl R., Herken R.;
 RT "The extracellular matrix proteins fibulin-1 and fibulin-2 in the
 RT early human embryo";
 RL Histochem. J. 26:109-116(1996).
 [13]
 RN INDUCTION.
 RP MEDLINE=96133928; PubMed=8552629;
 RA Clinton G.N., Rougeot C., Derancourt J., Roger P., Defrenne A.,
 RA Godyna S., Argaves W.S., Rochefort H.;
 RT "Estrogens increase the expression of fibulin-1, an extracellular
 RT matrix protein secreted by human ovarian cancer cells";
 RL Proc. Natl. Acad. Sci. U.S.A. 93:318-320(1996).
 [14]
 RN CALCUM, SELF-ASSOCIATION, AND FN1-BINDING SITES.
 RP PubMed=9278415;
 RA Tran H., VanDusen W.J., Argaves W.S.;
 RT "The self-association and fibronectin-binding sites of fibulin-1 map
 RT to calcium-binding epidermal growth factor-like domains";
 RL J. Biol. Chem. 272:22600-22606(1997).
 [15]
 RN ROLE IN TUMOR FORMATION AND INVASION.
 RP PubMed=9393974;
 RA Qing J., Maher V.X., Tran H., Argaves W.S., Dunstan R.W.,
 RA McCormick J.;
 RT "Suppression of anchorage-independent growth and matrix invasion and
 RT delayed tumor formation by elevated expression of fibulin-1D in human
 RT fibrosarcoma-derived cell lines";
 RL Oncogene 15:2159-2168(1997).
 [16]
 RN INDUCTION.
 RP PubMed=9811350;
 RA Roger P., Pujol P., Lucas A., Baidet P., Rochefort H.;
 RT "Increased immunostaining of fibulin-1, an estrogen-regulated protein
 RT in the stroma of human ovarian epithelial tumors";
 RL Am. J. Pathol. 153:1579-1588(1998).
 [17]
 RN ROLE IN TUMOR FORMATION AND INVASION.
 RP PubMed=9466571;
 RA Hayashida Y., Lucas A., Rougeot C., Godyna S., Argaves W.S.,
 RA Rochefort H.;
 RT "Estradiol and fibulin-1 inhibit motility of human ovarian- and
 RT breast-cancer cells induced by fibronectin";
 RL Int. J. Cancer 75:654-658(1998).
 [18]
 RN INTERACTION WITH NOV.
 RP PubMed=9927860;
 RA Perbal B., Martinier C., Sainson R., Werrin M., He B., Polzman B.;
 RT "The C-terminal domain of the regulatory protein NOV is sufficient to
 RT Query Match 10.1%; Score 567; DB 1; Length 703;
 Best Local Similarity 28.8%; Pred. No. 1.9e-30;
 Matches 159; Conservative 75; Mismatches 167; Indels 152; Gaps 29;
 Oy 44 EDVDECAQGLDDCHADALCONTFTSYK----SCKPGYQ-GEGRCEIDECGNELNGCC 98
 Db 2:5 EDVNECTGSHSCRLGESCINTVGSFRCDSSCGTYELTDSNCKDIDECESIH-NC 273
 Cy 99 VHP--CNPNGYRC-----TCDFGNLHAGNCLDVLDECLNNGCC--HTCVNWMGSY 150
 Db 274 LPDFICONTLGSFRCRFKLQCKSGFIQALG-NCIDINECLISAPCTIGTCTNTEGSS 332
 Cy 151 EC-----CCKGKGFLLSDNQTCHIRSEGLSCMNKDH-----GCSEHCKEAFRGSSVA 197
 Db 333 TCQKNVFNCGRGY-----HUNEETRCVQVDECAAPPMEPCGKGRVNSF-GSFR 381

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM: protein - protein search, using sw model

Run on: October 22, 2003, 11:16:54 ; Search time 88 Seconds
(without alignments)
1801.907 Million cell up

Title: US-09-747-371-2

perfect score:

Sequence: 1 MGVAGRNRPGAAWAVLLLL...PRSFIRLLRSKVSRFLRPYK 999

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters:

Minimum PG seq length: 0

Minimum	DB	seq	length: 0
Maximum	DB	seq	length: 200000000

Post-processing: Minimum Match 0%

Post-processing: Minimum March 0% Maximum March 100%

Maximum PACCH 100%
Listing first 45 summaries

Databas : A Geneset 19.Jun03:*

Database : *A. nidens* 1200n301

1:	/SIDS1/cgcdata/geneseq/geneseq-emb1/AA1980.DAT.*
2:	/SIDS1/cgcdata/geneseq/geneseq-emb1/AA1981.DAT.*
3:	/SIDS1/cgcdata/geneseq/geneseq-emb1/AA1982.DAT.*
4:	/SIDS1/cgcdata/geneseq/geneseq-emb1/AA1983.DAT.*
5:	/SIDS1/cgcdata/geneseq/geneseq-emb1/AA1984.DAT.*
6:	/SIDS1/cgcdata/geneseq/geneseq-emb1/AA1985.DAT.*
7:	/SIDS1/cgcdata/geneseq/geneseq-emb1/AA1986.DAT.*
8:	/SIDS1/cgcdata/geneseq/geneseq-emb1/AA1987.DAT.*
9:	/SIDS1/cgcdata/geneseq/geneseq-emb1/AA1988.DAT.*
10:	/SIDS1/cgcdata/geneseq/geneseq-emb1/AA1989.DAT.*
11:	/SIDS1/cgcdata/geneseq/geneseq-emb1/AA1990.DAT.*
12:	/SIDS1/cgcdata/geneseq/geneseq-emb1/AA1991.DAT.*
13:	/SIDS1/cgcdata/geneseq/geneseq-emb1/AA1992.DAT.*
14:	/SIDS1/cgcdata/geneseq/geneseq-emb1/AA1993.DAT.*
15:	/SIDS1/cgcdata/geneseq/geneseq-emb1/AA1994.DAT.*
16:	/SIDS1/cgcdata/geneseq/geneseq-emb1/AA1995.DAT.*
17:	/SIDS1/cgcdata/geneseq/geneseq-emb1/AA1996.DAT.*
18:	/SIDS1/cgcdata/geneseq/geneseq-emb1/AA1997.DAT.*
19:	/SIDS1/cgcdata/geneseq/geneseq-emb1/AA1998.DAT.*
20:	/SIDS1/cgcdata/geneseq/geneseq-emb1/AA1999.DAT.*
21:	/SIDS1/cgcdata/geneseq/geneseq-emb1/AA2000.DAT.*
22:	/SIDS1/cgcdata/geneseq/geneseq-emb1/AA2001.DAT.*
23:	/SIDS1/cgcdata/geneseq/geneseq-emb1/AA2002.DAT.*
24:	/SIDS1/cgcdata/geneseq/geneseq-emb1/AA2003.DAT.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	5605	100.0	999	23	ABP63129	Human polypeptide
2	5605	100.0	999	23	ABJ05567	Breast cancer-asso
3	5605	100.0	999	23	ABB0926	Human breast cancer
4	5605	100.0	999	23	ABG61893	Prostate cancer-asso
5	5605	100.0	999	24	ABR47407	Breast cancer asso
6	5605	100.0	999	24	ABJ19815	Androgen-independent
7	5581	99.6	997	24	ABJ37049	Human breast cancer
8	5428	96.8	964	24	ABP56757	Human EGFP protein
9	5079	90.6	919	24	AAO15642	Human extracellular

ALIGNMENTS

RESULT 1

ABP69329

IC ABP69329 standard; Protein; 999 AA.

AC ABP69329:

20-11AN-2003 (first entry)

XX
DE
Shirley M. [vent] e
CES P. CN . 76

Human; genome mapping; gene therapy; food supplement; virus; fungus;
cell; regenerative disorder; neurodegenerative disease; bacterial;
Parkinson's disease; Alzheimer's disease; autoimmune disease;
multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;
arthritis; cyclostatic; immunomodulator; nontropic; neuroprotective;
anti-Parkinsonian; antidiabetic; immunosuppressive; dermatological;
haemostatic; vulvectomy; fungicide; antibacterial; virucide; protozoacide;
antiarrhythmic.

xx
sc
os
Homo sapiens.

XX
PN
WC2C0270539-A2

2
 2
 2
 -
 6
 6
 6
 2-2
 .

[illegible]

XX
XX
XX

XX

XX

Mouse	BCG2 ortholog
Mouse	CEBP1 protein
Human	Epidermal growth factor receptor
Human	Protein of NOV8 8b
Human	MEGF/FIBRILL
Human	Novel human epider
Human	LP283 protein
Human	POLY7 protein
Human	LP283 mature
Amino acid	sequence
Human	LP283 splice
Human	LP283 protein
Human	extracellular
Novel	Human epider
Novel	acid sequence
Amino acid	sequence
Protein of NOV8	8a
Human	EGF-Related
Mouse	EGF-like 1 p
Human	LP283 splice
Human	POLY6 protein
Novel	human epider
Novel	human diagno
Amino acid	sequence
Primate LP283	Var
Human	breast-speci
Human	POLY5 protein
Novel	human epider
Human	MDMT polyPSP
Breast cancer	prob
Human	ovarian anti
Amino acid	sequence
Human	56719 amino
Human	LP283 splice
Amino acid	sequence
Human	CDNA SECID

PI Xue AJ, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M;
PI Wehrman T, Wang J, Wang D, Drmanac RT;
XX WPI; 2002-759812/82.
DR N-PSDB; AB211546.
XX
XX New polynucleotides comprising sequences assembled from expressed
PT sequence tags (ESTs), useful for treating cell-proliferative, or
PT neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or
PT platelet or coagulation disorders -
XX
XX Claim 9; SEQ ID NO 1376; 1012pp + Sequence Listing; English.
XX
XX The invention relates to an isolated polynucleotide (1) comprising a
CC nucleotide sequence selected from any of 948 sequences
CC (AB2111.9-AB211066) or their mature protein coding portion, active domain
CC coding protein or complementary sequences. The polynucleotides are useful
CC for identifying expressed genes or for physical mapping of human genome.
CC The encoded polypeptides (ABP6902-ABP6949) are useful as molecular
CC weight markers, as a food supplement, for generating antibodies, in
CC medical imaging, screening and diagnostic assays and for treating
CC cell-proliferative disorders (cancer), neurodegenerative diseases
CC (Parkinson's or Alzheimer's disease), autoimmune diseases (multiple
CC sclerosis, diabetes, lupus) genetic disorders, myeloid or lymphoid
CC disorders, platelet or coagulation disorders, wound, burns, incision,
CC ulcers, liver or lung fibrosis, infections (bacterial, viral, fungal,
CC parasitic), arthritis, etc.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 999 AA;

Query Match 100.0%; Score 5605; DB 23; Length 999;
Best Local Similarity 100.0%; Fred. No. 0;
Matches 999; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGAGNRGGAAMVALLLLPPLLLLAGAVPPGRAGAGPOEDVDECAOGLDCHADA 60
Db 1 MGAGNRGGAAMVALLLLPPLLLLAGAVPPGRAGAGPOEDVDECAOGLDCHADA 60
QY 61 LCQNTPTSKYCKCPGYQEGRQCEDIDECGNELNGGCVHDCINIPGNVRYCTCFDGMFLA 120
Db 61 LCQNTPTSKYCKCPGYQEGRQCEDIDECGNELNGGCVHDCINIPGNVRYCTCFDGMFLA 120
QY 121 HDGNCLDVECLNNGGQHTCVNMGSYECCKEGFELSPNQHCTCIHRSEEG;SOMNK 180
Db 121 HDGNCLDVECLNNGGQHTCVNMGSYECCKEGFELSPNQHCTCIHRSEEG;SOMNK 180
QY 181 DRGCSHICKEAPRGSVACERPGFELAKNQRDCILTCNHNGSGCQHSDDTADGPECSC 240
Db 181 DRGCSHICKEAPRGSVACERPGFELAKNQRDCILTCNHNGSGCQHSDDTADGPECSC 240
QY 241 PQYKHTDGRSLEREDTVLETSNTTSVWDGKRVKRLMETCAVNGGCDRTCKDT 300
Db 241 PQYKHTDGRSLEREDTVLETSNTTSVWDGKRVKRLMETCAVNGGCDRTCKDT 300
QY 301 STGVHSCVPVGTFLQDGTGKTDIDECQTRNGGCHDFCKNI VGSFDCGCKGFKLLTDEK 360
Db 301 STGVHSCVPVGTFLQDGTGKTDIDECQTRNGGCHDFCKNI VGSFDCGCKGFKLLTDEK 360
QY 361 SCQDVDECSLDRTCDHSCINHPGTAFACNRYGTYLGFTHCGDTNECSINNGGCGQCVN 420
Db 361 SCQDVDECSLDRTCDHSCINHPGTAFACNRYGTYLGFTHCGDTNECSINNGGCGQCVN 420
QY 421 TVGSYECQCHPGYKLVHNNKDCVEVWGLLPTSVSRVSLHCGKSGGCGCFLRCHSG;HL 480
Db 421 TVGSYECQCHPGYKLVHNNKDCVEVWGLLPTSVSRVSLHCGKSGGCGCFLRCHSG;HL 480
QY 481 SSVTVTIRTSVTFKLNKGC SLKNAELFPEGLRPALPEKHSSVKSFRVNLTCSSGQV 540
Db 481 SSVTVTIRTSVTFKLNKGC SLKNAELFPEGLRPALPEKHSSVKSFRVNLTCSSGQV 540

QY 541 PGAPRPSTPKEMFIVTEFELETNQKEVTASCDLSCLVAPTEKRLKA;RTLRKAVHREQ 600
Db 541 PGAPRPSTPKEMFIVTEFELETNQKEVTASCDLSCLVAPTEKRLKA;RTLRKAVHREQ 600
QY 601 FHLQLSGMNLDVAKKPPRTSERQAESCGVGQGHAEHQCVSCRAGTYDGGARERCILCPNG 660
Db 601 FHLQLSGMNLDVAKKPPRTSERQAESCGVGQGHAEHQCVSCRAGTYDGGARERCILCPNG 660
QY 661 TFQNEBQMTCEPCPRPGNSGA;KTPBANMSECGGLCOFGGEYSADGFAPQC;CALGTQ 720
Db 661 TFQNEBQMTCEPCPRPGNSGA;KTPBANMSECGGLCOFGGEYSADGFAPQC;CALGTQ 720
QY 721 PEAGRTSCFFCGGLATKHGATSCFCCETRVOCSPGHFVNTTHRCISCPVGTQPEFG 780
Db 721 PEAGRTSCFFCGGLATKHGATSCFCCETRVOCSPGHFVNTTHRCISCPVGTQPEFG 780
QY 781 KNCVSCPGNTTTFDGSNTITCKNRRCGGELGDTFGVIESPNYPNYPANTECTWTIN 840
Db 781 KNCVSCPGNTTTFDGSNTITCKNRRCGGELGDTFGVIESPNYPNYPANTECTWTIN 840
QY 841 PPKRRRLVVPFELFIEDDCQY;VMKTSNNSVTTETCQTYERPIAFTSRSKKLW 900
Db 841 PPKRRRLVVPFELFIEDDCQY;VMKTSNNSVTTETCQTYERPIAFTSRSKKLW 900
QY 901 QPKSNEGNSARGQVQVYVYTYDEDYDELIED;VRDGLVASENHQELKOKK;KALFDV 960
Db 901 QPKSNEGNSARGQVQVYVYTYDEDYDELIED;VRDGLVASENHQELKOKK;KALFDV 960
QY 961 LAHPQYKVTAGSEMEPRSPFIRLRSKVSFLRPYK 999
Db 961 LAHPQYKVTAGSEMEPRSPFIRLRSKVSFLRPYK 999

RESULT 2
AB05567
ID AB05567 standard; Protein: 999 AA.
XX AC AB05567;
XX DT 14-NOV-2002 (first entry)
XX DE Breast cancer-associated protein 32.
XX KW Breast cancer; breast cancer-associated gene sequence;
XX KW drug development; pharmacogenetics; biosensor development.
XX OS Unidentified.
XX FN W0200259377-A2.
XX PD 01-AUG-2002.
XX PF 24-JAN-2002; 2002WO-US02242.
XX PR 24-JAN-2001; 2001US-263965P.
XX PR 02-FEB-2001; 2001US-265928P.
XX PR 09-APR-2001; 2001US-282698P.
XX PR 09-APR-2001; 2001US-0829472.
XX PR 04-MAY-2001; 2001US-288590P.
XX PR 29-MAY-2001; 2001US-29443P.
XX (ECSB-) EOS BIOTECHNOLOGY INC.
XX PA Yack DH, Gish KC, Afar D;
XX PI WPI; 2002-583738/62.
XX DR N-PSDB; AB07724.
XX PT Detecting a breast cancer-associated transcript in a patient's cell.
XX PT useful for diagnosing breast cancer, comprises contacting a biological
XX PT sample with a polynucleotide that selectively hybridizes with breast
XX PT cancer nucleic acids

PS Disclosure; Page 375; 414pp; English.

XX The invention comprises a method of detecting a breast cancer-associated transcript in a cell from a patient. The method of the invention involves contacting a biological sample from the patient with a nucleotide that hybridises to one of the 69 breast cancer-associated gene sequences shown in the specification. The method of the invention is useful in the diagnosis or prognosis of breast cancer, and for detecting genes that are up or down-regulated in breast cancer cells. Genes identified by the method of the invention can be used in diagnostic purposes and also as targets for screening for therapeutic compounds that modulate breast cancer (e.g. hormones or antibodies). Identification of genes that are over or under expressed in breast cancer can additionally provide high-resolution, high-sensitivity datasets which can be used in the areas of diagnostics, therapeutics, drug development, pharmacogenetics, protein structure and biosensor development. Amino acid sequences AB05536 - AB05604 represent the proteins encoded by the 69 breast cancer-associated genes of the invention.

XX Sequence 999 AA;

Query Match 100.0%; Score 5605; DB 23; Length 999;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 999; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGVAAGNRGAAGAVALLLLPPLLLLAGAVPPGGRGAAGPQEDVDECAQGLDDCHADA 60
 DB 1 MGVAAGNRGAAGAVALLLLPPLLLLAGAVPPGGRGAAGPQEDVDECAQGLDDCHADA 60
 QY 61 LCQNTPTSKYCKPCYQGGQCEDECEGELNGGCVHDCNLIPGNYRCTCPGFMIA 120
 DB 61 LCQNTPTSKYCKPCYQGGQCEDECEGELNGGCVHDCNLIPGNYRCTCPGFMIA 120
 QY 121 HGHNCNLDVDECLNNGGQHTCVNVMGSEYCCCKEGFELSDNQHTCIHRSEGLSCMNK 180
 DB 121 HGHNCNLDVDECLNNGGQHTCVNVMGSEYCCCKEGFELSDNQHTCIHRSEGLSCMNK 180
 QY 181 DHGCSHCKEAPRGSVACRCRPFELAKNQRDCLTCNHNNGGCGQHSDDTDAGPECSC 240
 DB 181 DHGCSHCKEAPRGSVACRCRPFELAKNQRDCLTCNHNNGGCGQHSDDTDAGPECSC 240
 QY 241 PQYKMTDGRSCLEEDTVEJTESNTTSVVDGDKVRLMETCAVNNGGCDTCKDT 300
 DB 241 PQYKMTDGRSCLEEDTVEJTESNTTSVVDGDKVRLMETCAVNNGGCDTCKDT 300
 QY 301 STGVHSCRPVFTLQDGGTKCDIDECQTRNGSCDHFCKNIVGSEDCGCKGFKLLTDEK 360
 DB 301 STGVHSCRPVFTLQDGGTKCDIDECQTRNGSCDHFCKNIVGSEDCGCKGFKLLTDEK 360
 QY 361 SCQVDECSLDRDTCDSHCINHPGTACACNRGYTLVGFTHCGDTNECSINNGGCGQCVN 420
 DB 361 SCQVDECSLDRDTCDSHCINHPGTACACNRGYTLVGFTHCGDTNECSINNGGCGQCVN 420
 QY 421 TVGSYECQCHPGYKHLNKKDCVEVKGLLPTSVPRVSLHCGKSGGDCFLRCHSGIHL 480
 DB 421 TVGSYECQCHPGYKHLNKKDCVEVKGLLPTSVPRVSLHCGKSGGDCFLRCHSGIHL 480
 QY 481 SSDVTITRTSVTFKLNKCSLKNABLPPEGLRPALEKHSVSKESFRVNLTCSSGQV 540
 DB 481 SSDVTITRTSVTFKLNKCSLKNABLPPEGLRPALEKHSVSKESFRVNLTCSSGQV 540
 QY 541 PGAPGRSPKPNFIVFELETNQKVTASCDLSICIVKTEKRLKAKIIRTKAVHREQ 600
 DB 541 PGAPGRSPKPNFIVFELETNQKVTASCDLSICIVKTEKRLKAKIIRTKAVHREQ 600
 QY 601 FHLQSGMLDVAKKPPRTSROAESCGVQGHAEQVSCAGTYDGCARECILLCPNG 660
 DB 601 FHLQSGMLDVAKKPPRTSROAESCGVQGHAEQVSCAGTYDGCARECILLCPNG 660
 QY 661 TFQNEEGQMTCEPCPRPGNSGALKTPEAWNMSECGGLCPGYSADGAPACOLCALGTFQ 720
 DB 661 TFQNEEGQMTCEPCPRPGNSGALKTPEAWNMSECGGLCPGYSADGAPACOLCALGTFQ 720

QY 721 PEAGRTSCFPCCGGLATKQAGATSFQCCETRVQCSQGHFVNTTTHRCIRCPVGTQYQPEFG 780
 DB 721 PEAGRTSCFPCCGGLATKQAGATSFQCCETRVQCSQGHFVNTTTHRCIRCPVGTQYQPEFG 780
 QY 781 KKNVCSPGNTTDFDGSNTNCTCKNRCGGEJGDFGTGYESPNYPKNYPANTECTWTIN 840
 DB 781 KKNVCSPGNTTDFDGSNTNCTCKNRCGGEJGDFGTGYESPNYPKNYPANTECTWTIN 840
 QY 841 PPKRRILLIVPRIFPIEDDCGYLVWRKTSNSNSVTTTCTYERPAFTSRSKKLW 900
 DB 841 PPKRRILLIVPRIFPIEDDCGYLVWRKTSNSNSVTTTCTYERPAFTSRSKKLW 900
 QY 901 IQPKSNEGNSARGQFQVYVYDGYGELIEDIVDRGLYASENHQELDKKLIKALJFV 960
 DB 901 IQPKSNEGNSARGQFQVYVYDGYGELIEDIVDRGLYASENHQELDKKLIKALJFV 960
 QY 961 LAHPONYFKYTAQESREYFPRSPFRLLRKSVSRFLRPPYK 999
 DB 961 LAHPONYFKYTAQESREYFPRSPFRLLRKSVSRFLRPPYK 999
 RESULT 3
 ASB80926
 ID ASB80926 standard; Protein: 999 AA.
 XX ASB80926;
 AC
 XX
 DT 08-OCT-2002 (first entry)
 XX
 DE Human breast cancer modulating protein BCO2.
 XX
 DE Breast cancer modulating protein; BCMP; BCO2; cytostatic; human;
 KW immunostimulant; antisense therapy; gene therapy; vaccine.
 XX
 OS Homo sapiens.
 XX
 XX WO200255988-A2.
 XX 18-JUL-2002.
 XX
 XX 21-DEC-2001; 2001WO-US50548.
 XX
 XX 21-DEC-2000; 2000US-0747371.
 XX
 XX (ECSB-); EOS BIOTECHNOLOGY INC.
 PA
 Gish KC, Mack D;
 XX
 XX WPT; 2002-566749/60.
 DB N-PSDB; ARN86363.
 XX
 PT Diagnosing breast cancer comprises comparing the gene expression of a
 breast cancer modulating protein in a test breast tissue and unaffected
 breast tissue of the same or a different patient
 XX
 PS Disclosure; Fig 2; 68pp; English.
 XX
 XX The invention relates to diagnosing breast cancer that involves
 determining the expression of a gene encoding breast cancer modulating
 protein (BCO2) or a fragment of it in a breast tissue of an individual,
 and comparing the expression of the genes from a second normal breast
 tissue from the individual or a second unaffected individual, where a
 difference in expression indicates the individual has breast cancer. The
 method is used for diagnosing breast cancer. Other new methods are
 provided for screening drug candidates, screening for bioactive agents,
 and evaluating the effect of a candidate, screening for bioactive agents,
 to BCO2 can be used to inhibit the bioactivity of BCO2 or neutralise it's
 effect and can be used to treat breast cancer. The antibody can also be
 used to localise a therapeutic group to breast cancer tissue, which can
 be used to treat breast cancer. Antisense molecules are used to inhibit
 breast cancer in a cell. BCO2 protein and polynucleotides are used to
 elicit an immune response in an individual. BCO2 is also used to
 determining the prognosis of an individual with breast cancer. The

CC present sequence represents the human BCO2 protein.

XX	Sequence	999 AA;	Query Match	100.0%;	Score 5605;	DB 23;	Length 999;
XX	Sequence	999 AA;	Best Local Similarity	100.0%;	Pred. No. 0;		
XX	Sequence	999 AA;	Matches 999;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	MGVAGNRPGAAWAVLLLLLLLLLLLLLLLLLLLLLAGAVPPGRGAAGPOEDVDECAQGLDDCHADA	60				
Dy	1	MGVAGNRPGAAWAVLLLLLLLLLLLLLLLLLLLLLAGAVPPGRGAAGPOEDVDECAQGLDDCHADA	60				
Qy	61	LCQNTPTSYKCSCKPGVGGGROCEDIDECNGELNGSCVHDCUNIPGNVRCCTCFGEVLA	120				
Dy	61	LCQNTPTSYKCSCKPGVGGGROCEDIDECNGELNGSCVHDCUNIPGNVRCCTCFGEVLA	120				
Qy	121	HDGHNCLDVECLJENGGCQHTCVNMGSYECCKEGFFLSNQHTCIHRSSEGLSCNMK	180				
Dy	121	HDGHNCLDVECLJENGGCQHTCVNMGSYECCKEGFFLSNQHTCIHRSSEGLSCNMK	180				
Qy	181	DHGCISHICREAPRGVACBPCPGFELAKNQRDCILTCNHNGGCGQSCDDTADGPECSCH	240				
Dy	181	DHGCISHICREAPRGVACBPCPGFELAKNQRDCILTCNHNGGCGQSCDDTADGPECSCH	240				
Qy	241	POYKMTDGRSLEREDTVLEVTESNTTSVVDGDKVRKRLMETCAVNGGCDRTCKDT	300				
Dy	241	POYKMTDGRSLEREDTVLEVTESNTTSVVDGDKVRKRLMETCAVNGGCDRTCKDT	300				
Qy	301	STGVHSCVPVFTLQDGKTKCDIDECQTRNGGCDHFCKNIVGSPFCGCKGFKLLTDEK	360				
Dy	301	STGVHSCVPVFTLQDGKTKCDIDECQTRNGGCDHFCKNIVGSPFCGCKGFKLLTDEK	360				
Qy	361	SCQDVDECSLDRTCDHSCINRGTFACACNRYGTYGFTHCGDTNRCSTNNGCQGVN	420				
Dy	361	SCQDVDECSLDRTCDHSCINRGTFACACNRYGTYGFTHCGDTNRCSTNNGCQGVN	420				
Qy	421	TGVSQCQCHPGYKJLHWNKKDCEVVEKGLLPTSVSRVSLHCGKGGGCGFLRCHSGIHL	480				
Dy	421	TGVSQCQCHPGYKJLHWNKKDCEVVEKGLLPTSVSRVSLHCGKGGGCGFLRCHSGIHL	480				
Qy	481	SSDVTIRTSVTKLNBKGLSKNAELFPEGLRALPEKHSVVKESFRVYNTTCSGKGV	540				
Dy	481	SSDVTIRTSVTKLNBKGLSKNAELFPEGLRALPEKHSVVKESFRVYNTTCSGKGV	540				
Qy	541	PGAPGRSTPKEMFTTVEPELETNQEVTASCDLSICIVKRTKRLKA-RTLKKAHREQ	600				
Dy	541	PGAPGRSTPKEMFTTVEPELETNQEVTASCDLSICIVKRTKRLKA-RTLKKAHREQ	600				
Qy	601	FHLQLGSMNLIDVAKKPTTSERQAESCQGVQGHAEQCVSCRAGTYDGAERECILCPNG	660				
Dy	601	FHLQLGSMNLIDVAKKPTTSERQAESCQGVQGHAEQCVSCRAGTYDGAERECILCPNG	660				
Qy	661	TFONEGQMTCEPCPRGNSGALKTPAANKMSCCGLCPGGEVSADGAPACOLCALGTTC	720				
Dy	661	TFONEGQMTCEPCPRGNSGALKTPAANKMSCCGLCPGGEVSADGAPACOLCALGTTC	720				
Qy	721	PEAGRTSCFPGGGLATKHKGATSFQDCETRVQCSQGHFVNTTTHRCIRCPVGTQPEFS	780				
Dy	721	PEAGRTSCFPGGGLATKHKGATSFQDCETRVQCSQGHFVNTTTHRCIRCPVGTQPEFS	780				
Qy	781	KNCVSCPGNTTDFDGSNTITCKNRRCGGELGDFGTGIESPNYPGNTPANTECTWTN	840				
Dy	781	KNCVSCPGNTTDFDGSNTITCKNRRCGGELGDFGTGIESPNYPGNTPANTECTWTN	840				
Qy	841	P2PKBRLLIVVPELFLP2EDCDGDLVWRKTSNNSUTTYETCQYERPIAFTSRKLLN	900				
Dy	841	P2PKBRLLIVVPELFLP2EDCDGDLVWRKTSNNSUTTYETCQYERPIAFTSRKLLN	900				
Qy	901	IQFKNSEGNARGQVPPVYTYDEYQELIEDIVRDGRLVASENEQELDKKLLKALFDV	960				
Dy	901	IQFKNSEGNARGQVPPVYTYDEYQELIEDIVRDGRLVASENEQELDKKLLKALFDV	960				
Qy	961	LAHPQNYFYKTAQBSREMFPRSFIRLLRSKVSFLRPHYK	999				

Db 961 LAHPQNYFYKTAQBSREMFPRSFIRLLRSKVSFLRPHYK 999

RESULT 4

ABG61893	ID	ABG61893	standard; Protein; 999 AA.
XX	AC	ABG61893;	
XX	DT	15-AUG-2002	(first entry)
XX	DE	Prostate cancer-associated protein #94.	
XX	KW	Prostate cancer; prostate tumour tissue; human; mammary; cystostatic.	
XX	OS	Mammalia.	
XX	PN	WO200230268-A2.	
XX	PD	19-APR-2002.	
XX	PF	12-OCT-2001; 2001WC-US32045.	
XX	PR	13-OCT-2000; 2000US-0687576.	
XX	PR	08-DEC-2000; 2000US-0733388.	
XX	PR	08-DEC-2000; 2000US-0733342.	
XX	PR	24-JAN-2001; 2001US-263957P.	
XX	PR	16-MAR-2001; 2001US-276797P.	
XX	PR	16-MAR-2001; 2001US-276888P.	
XX	PR	06-APR-2001; 2001US-281922P.	
XX	PR	24-APR-2001; 2001US-286214P.	
XX	PR	30-APR-2001; 2001US-0847046.	
XX	PR	04-MAY-2001; 2001US-286599P.	
XX	PA	(E05B-) EOS BIOTECHNOLOGY INC.	
XX	PI	Gish KC, Mack DH, Wilson KE, Afar D, Hevezi P;	
XX	DR	WPI; 2002-471335/50.	
XX	DR	N-PSDB; ABK92209.	
XX	PT	Detecting a prostate cancer-associated transcript in a cell in a	
XX	PT	patient, useful for diagnosing prostate cancer (PC) or screening	
XX	PT	modulators of PC, by determining if prostate cancer-associated genes	
XX	PT	are expressed in a prostate tissue	
XX	PS	Claim 27; Page 379; 43pp; English.	
XX	CC	The present invention relates to methods of detecting a prostate	
XX	CC	cancer-associated transcript in a cell from a patient. The method	
XX	CC	comprises contacting a biological sample from the patient with	
XX	CC	prostate cancer-associated polynucleotides (designated PC genes) that	
XX	CC	selectively hybridize to a sequence that is at least 80% identical	
XX	CC	to them. The prostate cancer-associated polynucleotide sequences	
XX	CC	are differentially expressed in prostate tumour tissue or in	
XX	CC	prostate cancer and are derived from the tissues of various	
XX	CC	organisms such as humans or other mammals (e.g. mice, sheep and dogs).	
XX	CC	The methods of the invention are useful for diagnosing and treating	
XX	CC	prostate cancer in mammals. The prostate cancer-associated genes are	
XX	CC	useful for diagnosing or treating prostate cancer, as well as for	
XX	CC	identifying modulators of prostate cancer or agents that inhibit	
XX	CC	prostate cancer. The nucleic acid sequences are particularly useful	
XX	CC	in gene therapy, as a vaccine or in antisense applications.	
XX	CC	ABG61800-ABG61944 represent prostate cancer-associated proteins.	
XX	SQ	Sequence 999 AA;	

Query Match 100.0%; Score 5605; DB 23; Length 999;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 999; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy : MGVAGNRPGAAWAVLLLLLLLLLLLLLAGAVPPGRGAAGPOEDVDECAQGLDDCHADA 60

QY 241 POYKHTDGRSCLEREDTVLEVTESNTTSVVDGKSKVRLLMETCAVNNGGCDRTCKOT 300
 DB 241 POYKHTDGRSCLEREDTVLEVTESNTTSVVDGKSKVRLLMETCAVNNGGCDRTCKOT 300
 QY 301 STGVHCSCPVGTFLQDGTCKDIDECOTRNGGCHDFCKNTVGSFDCGCKGFKLLTDEK 360
 DB 301 STGVHCSCPVGTFLQDGTCKDIDECOTRNGGCHDFCKNTVGSFDCGCKGFKLLTDEK 360
 QY 361 SCQVDECSLDRTCDHSCINHPGTTFACACNRGYTLVGFTHCGDTNECSINNGCQVCVN 420
 DB 361 SCQVDECSLDRTCDHSCINHPGTTFACACNRGYTLVGFTHCGDTNECSINNGCQVCVN 420
 QY 421 TVGSVECOCHPGYKLVHMKDCEVKGLLPTSVSPVSLHCKSGGGGCFRCHSGIHL 480
 DB 421 TVGSVECOCHPGYKLVHMKDCEVKGLLPTSVSPVSLHCKSGGGGCFRCHSGIHL 480
 QY 481 SSVDTTIRTSVTFKLNKCSLKNAELFPEGRLPALPEKHSVSKSFYVNLTCSSGQV 540
 DB 481 SSVDTTIRTSVTFKLNKCSLKNAELFPEGRLPALPEKHSVSKSFYVNLTCSSGQV 540
 QY 541 PGAPGRSPCKEMFTVEPELTNQKEVTASCDLSCIIVKRTKRLKAIIRLKAHREQ 600
 DB 541 PGAPGRSPCKEMFTVEPELTNQKEVTASCDLSCIIVKRTKRLKAIIRLKAHREQ 600
 QY 601 FHLQSGMNLVAKPPRTSERQAESCQVGGQHAENQCVSCBAGTYDGAERCLCPNG 660
 DB 601 FHLQSGMNLVAKPPRTSERQAESCQVGGQHAENQCVSCBAGTYDGAERCLCPNG 660
 QY 661 TFCNEGQMTCEPCRPNGSGALKTPKPEANMSECGGCPGYSADGAPCOLALGTFO 720
 DB 661 TFCNEGQMTCEPCRPNGSGALKTPKPEANMSECGGCPGYSADGAPCOLALGTFO 720
 QY 721 PEAGRTSCPPCGGLATKHQATSFQDCETRVQCSPGHFYNTTTHRCRCPVGTVPQPEFG 780
 DB 721 PEAGRTSCPPCGGLATKHQATSFQDCETRVQCSPGHFYNTTTHRCRCPVGTVPQPEFG 780
 QY 781 KNNVSCPGNTTTFDGSNTITQKNRCGGELGDTGVIESPNYGPVNPANTECTWTIN 840
 DB 781 KNNVSCPGNTTTFDGSNTITQKNRCGGELGDTGVIESPNYGPVNPANTECTWTIN 840
 QY 841 PPPKRELLIIVPEIFLPIEDDCGDLVLRKTSNNSSVTTTETCTYERPIAFTSRSKLM 900
 DB 841 PPPKRELLIIVPEIFLPIEDDCGDLVLRKTSNNSSVTTTETCTYERPIAFTSRSKLM 900
 QY 901 IOFKNEGNSARGFQVYVYDDEYQELIEDIVROGRLYASENHQELIKKLIKALFDV 960
 DB 901 IOFKNEGNSARGFQVYVYDDEYQELIEDIVROGRLYASENHQELIKKLIKALFDV 960
 QY 961 LAHPQNYFYKTAGEGEMFPRFIRLLRSKVSRLRPYK 999
 DB 961 LAHPQNYFYKTAGEGEMFPRFIRLLRSKVSRLRPYK 999

RESULT 6
 AB:19815
 ID ABJ19815 standard; Protein; 999 AA.
 XX
 AC ABJ19815;
 XX
 DT 10-APR-2003 (first entry)
 XX
 DE Androgen-independent prostate cancer-related protein - SEQ ID NO 42.
 XX
 KW Androgen-independent cancer; androgen ablation therapy; prostate cancer;
 KW androgen-dependent prostate cancer; prostate cancer.
 XX
 CS Unidentified.
 XX
 FN WO200298358-A2.
 XX
 PD 12-DEC-2002.
 XX

PF 04-JUN-2002; 2002WO-US17594.
 XX 04-JUN-2001; 2001US-2959; 7P.
 PR 13-NOV-2001; 2001US-353666P.
 PR 29-MAR-2002; 2002US-368689P.
 PR 12-APR-2002; 2002US-372246P.
 PR 31-MAY-2002; 2002US-0160233.
 XX (ECSS-) EOS BIOTECHNOLOGY INC.
 XX Afar DEH, Agus C, Mack DH;
 PI WPI; 2003-148602/24.
 DB
 PT Detecting an androgen-independent prostate cancer cell in a sample or
 PT diagnosing androgen-dependent prostate cancer, by determining the
 PT presence or absence of genes whose expressions are up- or
 PT down-regulated.
 XX
 PS Claim 1; Page 199; 210pp; English.
 XX
 CC The invention comprises a method for detecting an androgen-independent
 CC cancer cell in a sample from a patient who has undergone androgen
 CC ablation therapy. The method involves determining the presence or absence
 CC of nucleic acids that are either up-regulated or down-regulated in
 CC prostate cancer. The method is useful for detecting an androgen-
 CC independent prostate cancer cell in a sample from a patient who has
 CC undergone androgen ablation therapy. The method is particularly useful
 CC for diagnosing androgen-dependent prostate cancer, prostate cancer
 CC undergoing androgen withdrawal, or androgen-independent prostate cancer.
 CC The present amino acid sequence represents a protein which is encoded by
 CC a gene that is either up-regulated or down-regulated in prostate cancer.
 XX
 SQ Sequence 999 AA;

Query Match: 100.0%; Score 5605; DB 24; Length 999;
 Best Local Similarity 100.0%; Pred. NC. C;
 Matches 999; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MSVAGNRPGAAWAVJLLLLPPLLACAVPPGRAGAPCEDYDECAQGLDCHADA 60
 DB 1 MSVAGNRPGAAWAVJLLLLPPLLACAVPPGRAGAPCEDYDECAQGLDCHADA 60
 QY 61 LCQNTPTSYKCSKPGYQGGROCEIDECGNEINGCVHDCJNIPQNYRCTCFDGMJA 120
 DB 61 LCQNTPTSYKCSKPGYQGGROCEIDECGNEINGCVHDCJNIPQNYRCTCFDGMJA 120
 QY 121 HDGHNCLDVEDELENNGGCQHTCVNVMGSVECCKEGFFLSDNQHTCIHRSEGLSCMK 180
 DB 121 HDGHNCLDVEDELENNGGCQHTCVNVMGSVECCKEGFFLSDNQHTCIHRSEGLSCMK 180
 QY 181 DGGSHICKAPRGSVACBPCGFELAKNORDILTCNHGNSGCCSDTAGPSCSCH 240
 DB 181 DGGSHICKAPRGSVACBPCGFELAKNORDILTCNHGNSGCCSDTAGPSCSCH 240
 QY 241 PQYKHTDGRSCLEREDTVLEVTESNTTSVVDGKSKVRLLMETCAVNNGGCDRTCKOT 300
 DB 241 PQYKHTDGRSCLEREDTVLEVTESNTTSVVDGKSKVRLLMETCAVNNGGCDRTCKOT 300
 QY 301 STGVHCSCPVGTFLQDGTCKDIDECQTRNGGCHDFCKNTVGSFDCGCKGFKLLTDEK 360
 DB 301 STGVHCSCPVGTFLQDGTCKDIDECQTRNGGCHDFCKNTVGSFDCGCKGFKLLTDEK 360
 QY 361 SCQVDECSLDRTCDHSCINHPGTTFACACNRGYTLVGFTHCGDTNECSINNGCQVCVN 420
 DB 361 SCQVDECSLDRTCDHSCINHPGTTFACACNRGYTLVGFTHCGDTNECSINNGCQVCVN 420
 QY 421 TVGSVECOCHPGYKLVHMKDCEVKGLLPTSVSPVSLHCKSGGGGCFRCHSGIHL 480
 DB 421 TVGSVECOCHPGYKLVHMKDCEVKGLLPTSVSPVSLHCKSGGGGCFRCHSGIHL 480
 QY 481 SSVDTTIRTSVTFKLNKCSLKNAELFPEGRLPALPEKHSVSKSFYVNLTCSSGQV 540
 DB 481 SSVDTTIRTSVTFKLNKCSLKNAELFPEGRLPALPEKHSVSKSFYVNLTCSSGQV 540

Db 481 SSDVTIRTSTVTKLNEGKSLKNAELFPEGLRALPEKHSSVKESFRYNLTCCSGKQV 540
 Qy 541 PGAPRSTPKEMFITVEFELETNQKEVTASCDLSICIVXRTKRLKAIKRTLRKAVHREQ 600
 Db 541 PGAPRSTPKEMFITVEFELETNQKEVTASCDLSICIVXRTKRLKAIKRTLRKAVHREQ 600
 Qy 601 FHLQSGMNLDAKKPPRTSERQAESCGVGCGGHAENQCVSCRASTYDYGARECILCPNG 660
 Db 601 FHLQSGMNLDAKKPPRTSERQAESCGVGCGGHAENQCVSCRASTYDYGARECILCPNG 660
 Qy 661 TFQNEEGOMTCEPCPRPGNSGALKTPAAMNMSCGGLCQPGESADGAPCOLCALGTQ 720
 Db 661 TFQNEEGOMTCEPCPRPGNSGALKTPAAMNMSCGGLCQPGESADGAPCOLCALGTQ 720
 Qy 721 PEAGRTSCPCGGGLATKHGATSFQDCETRVQCSFGHFYNTTTHRCIRCPVGTQPEFG 780
 Db 721 PEAGRTSCPCGGGLATKHGATSFQDCETRVQCSFGHFYNTTTHRCIRCPVGTQPEFG 780
 Qy 781 KNNVSCPGNTTDFDGSNTITQCKNRRCGGELGDTGYIESPNYSGNPANTECTWTIN 840
 Db 781 KNNVSCPGNTTDFDGSNTITQCKNRRCGGELGDTGYIESPNYSGNPANTECTWTIN 840
 Qy 841 PPPKRRILIVVPEIFLPIEDDCGYLVMKRTSSNSVTYETCQTYERPIAFTSRSKLW 900
 Db 841 PPPKRRILIVVPEIFLPIEDDCGYLVMKRTSSNSVTYETCQTYERPIAFTSRSKLW 900
 Qy 901 IQPKSNEGNSARGQVPPVYVYDQYQELIEDIVRDGRLYASENHQELKOKLIKALFDV 960
 Db 901 IQPKSNEGNSARGQVPPVYVYDQYQELIEDIVRDGRLYASENHQELKOKLIKALFDV 960
 Qy 961 LAHPQNYFKYTAQBSREMPFRSIRLLRSKVSRLRPYK 999
 Db 961 LAHPQNYFKYTAQBSREMPFRSIRLLRSKVSRLRPYK 999

RESULT 7

ID ABJ37049 standard; Protein; 997 AA.
 XX AC ABJ37049;
 D7 D7 01-MAY-2003 (first entry)
 XX DE Human: breast cancer / ovarian cancer related protein #25.
 DE Human; cytostatic; breast cancer; ovarian cancer.
 XX OS Homo sapiens.
 XX PN WO2003000012-A2.
 XX PD 03-JAN-2003.
 XX PF 21-JUN-2002; 2002WO-US19773.
 XX PR 21-JUN-2001; 2001US-300159P.
 XX PR 27-JUN-2001; 2001US-301351P.
 XX (MILL-) MILLENNIUM PHARM INC.
 PA Veiby OP;
 PI WP2; 2003-267848/26.
 DR N-PSDB; ABT3:918.
 PT Determining the presence of breast cancer in an individual, involves
 PT using specific polynucleotide markers -
 XX Disclosure; Page 141-143; 233pp; English.
 XX The invention comprises a method for assessing whether a patient is
 CC afflicted with breast cancer or ovarian cancer. The method involves the
 CC use of specific DNA markers. The method of the invention is useful in the

CC detection and treatment of ovarian and breast cancer. Amino acid
 CC sequences ABJ37049 - ABJ37049 represent human breast/ovarian cancer-
 CC related proteins.
 XX Sequence 997 AA:
 Qy 1 MGVAGRNRPAGAAVALLLLPPLLAGAVPPGRGAAGPQEDVDECAQGLDCHADA 60
 Db 1 MGVAGRNRPAGAAVALLLLPPLLAGAVPPGRGAAGPQEDVDECAQGLDCHADA 58
 Qy 61 LCQNTPTSYKSCAPGYQGEGRQCEDIDECNENLNGCVHDCNLIPGNYRCTCDGFMLA 120
 Db 59 LCQNTPTSYKSCAPGYQGEGRQCEDIDECNENLNGCVHDCNLIPGNYRCTCDGFMLA 118
 Qy 121 HDGNCILDVDCLENNGGCQHTCVNMVSGSYECCKEGFFLSNCHTCIHRSEBGLSCNKK 180
 Db 119 HDGNCILDVDCLENNGGCQHTCVNMVSGSYECCKEGFFLSNCHTCIHRSEBGLSCNKK 178
 Qy 181 DHGSHICKAEPGRSVACERPGFELAKNQRDCILTCNHGNGGCGCHSCDDADGPECSCH 240
 Db 179 DHGSHICKAEPGRSVACERPGFELAKNQRDCILTCNHGNGGCGCHSCDDADGPECSCH 238
 Qy 241 POYKMHDTGRSLCEREDTVLEVTESNTTSVVDGDKRVKRRLLMETCAVNNSGCORTCKOT 300
 Db 239 POYKMHDTGRSLCEREDTVLEVTESNTTSVVDGDKRVKRRLLMETCAVNNSGCORTCKOT 298
 Qy 301 STGVHSCSPVGTFLQDCKTKDIDECQTRNGGCDHFCKNIVGSPDCCKKGFLLTDEK 360
 Db 299 STGVHSCSPVGTFLQDCKTKDIDECQTRNGGCDHFCKNIVGSPDCCKKGFLLTDEK 358
 Qy 361 SCCQVDECSLDRTCDHSCINHPGTACACNRYTLYGFTHCGDTNECSINNGGCGQVQVW 420
 Db 359 SCCQVDECSLDRTCDHSCINHPGTACACNRYTLYGFTHCGDTNECSINNGGCGQVQVW 418
 Qy 421 TVGSYECQCHPEGYKLHMKKDCVEVKGLLPTSVSPRVSLHCKSGGGGDCFLRCHSGIHL 480
 Db 419 TVGSYECQCHPEGYKLHMKKDCVEVKGLLPTSVSPRVSLHCKSGGGGDCFLRCHSGIHL 478
 Qy 481 SSDVTIRTSTVTKLNEGKSLKNAELFPEGLRALPEKHSSVKESFRYNLTCCSGKQV 540
 Db 479 SSDVTIRTSTVTKLNEGKSLKNAELFPEGLRALPEKHSSVKESFRYNLTCCSGKQV 538
 Qy 541 PGAPRSTPKEMFITVEFELETNQKEVTASCDLSICIVXRTKRLKAIKRTLRKAVHREQ 600
 Db 539 PGAPRSTPKEMFITVEFELETNQKEVTASCDLSICIVXRTKRLKAIKRTLRKAVHREQ 598
 Qy 601 FHLQSGMNLDAKKPPRTSERQAESCGVGCGGHAENQCVSCRASTYDYGARECILCPNG 660
 Db 599 FHLQSGMNLDAKKPPRTSERQAESCGVGCGGHAENQCVSCRASTYDYGARECILCPNG 658
 Qy 661 TFQNEEGOMTCEPCPRPGNSGALKTPAAMNMSCGGLCQPGESADGAPCOLCALGTQ 720
 Db 659 TFQNEEGOMTCEPCPRPGNSGALKTPAAMNMSCGGLCQPGESADGAPCOLCALGTQ 718
 Qy 721 PEAGRTSCPCGGGLATKHGATSFQDCETRVQCSFGHFYNTTTHRCIRCPVGTQPEFG 780
 Db 719 PEAGRTSCPCGGGLATKHGATSFQDCETRVQCSFGHFYNTTTHRCIRCPVGTQPEFG 778
 Qy 781 KNNVSCPGNTTDFDGSNTITQCKNRRCGGELGDTGYIESPNYSGNPANTECTWTIN 840
 Db 779 KNNVSCPGNTTDFDGSNTITQCKNRRCGGELGDTGYIESPNYSGNPANTECTWTIN 838
 Qy 841 PPPKRRILIVVPEIFLPIEDDCGYLVMKRTSSNSVTYETCQTYERPIAFTSRSKLW 900
 Db 839 PPPKRRILIVVPEIFLPIEDDCGYLVMKRTSSNSVTYETCQTYERPIAFTSRSKLW 898
 Qy 901 IQPKSNEGNSARGQVPPVYVYDQYQELIEDIVRDGRLYASENHQELKOKLIKALFDV 960
 Db 899 IQPKSNEGNSARGQVPPVYVYDQYQELIEDIVRDGRLYASENHQELKOKLIKALFDV 958

QY 961 LAHPQNYFYKTAQESREMPRPSFIRLLRSKVSRLPYK 999
 DB 959 LAHPQNYFYKTAQESREMPRPSFIRLLRSKVSRLPYK 997

RESULT 8
 ABP56757
 ID ABP56757 standard; Protein; 964 AA.
 XX ASP56757;
 XX 31-MAR-2003 (first entry)
 XX Human CEGP1 protein SEQ ID NO:5.
 KW Human; secreted protein; epidermal growth factor protein subfamily;
 KW epidermal growth factor; EGF; gene therapy; therapeutic; drug screening;
 KW biological activity; immune response.
 XX Homo sapiens.
 XX M02002101080-A2.
 XX 19-DEC-2002.
 XX 07-MAY-2002; 2002MO-US22278.
 XX 16-MAY-2001; 2001US-0855824.
 XX (PEKE) PE CORP NY.
 XX (GONG/) GONG F.
 XX (DFRA/) DI FRANCESCO V.
 XX Gong F, Di Francesco V, Yan C, Beasley EM;
 XX WPI: 2003-167407/16.
 XX New isolated human secreted peptides, useful for diagnosing or treating
 PT a disease characterized by an absence of, inappropriate or unwanted
 PT expression of the secreted protein, and in drug screening assays -
 XX Disclosure; Fig 2E-F; 185pp; English.

XX The present sequence encodes a human secreted peptide (I) related to
 CC the epidermal growth factor (EGF) protein subfamily. (I) can be used
 CC in gene therapy. (I) and the nucleic acid molecules encoding (I) can
 CC be used as models for the development of human therapeutic targets, to
 CC aid in the identification of therapeutic proteins, and to serve as
 CC targets for the development of human therapeutic agents. (I) may be
 CC used in drug screening assays, in assays to determine the biological
 CC activity of the protein, to raise antibodies or to elicit another immune
 CC response, as a reagent in assays designed to quantitatively determine
 CC levels of the protein in biological fluids, or as markers for tissues
 CC in which the corresponding protein is preferentially expressed. (I) can
 CC also be used for diagnosing or treating a disease or disorder
 CC characterised by an absence of, inappropriate or unwanted expression of
 CC the protein. (I) is located to human chromosome 22. The present sequence
 CC represents human CEGP1 protein which is given in comparison with (I)
 CC in the exemplification of the present invention.

XX Query Match 96.8%; Score 5428; DB 24; Length 964;
 XX Best Local Similarity 100.0%; Pred. NO. 0;
 XX Matches 964; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 36 RGRAAGPQEDVDECAQGLDDCHADALCONPTSYKSCXGVOGEGRQCEDIDECGNELN 95
 DB 1 RGRAAGPQEDVDECAQGLDDCHADALCONPTSYKSCXGVOGEGRQCEDIDECGNELN 60

QY 96 GGCYHDCNLIPGNYRCTCFDGFMLAHGHNCLDVECLNNGGCQCHTCVNMVMSYECCCK 155
 XX
 XX Human extracellular messenger protein #2.

DB 61 GGCYHDCNLIPGNYRCTCFDGFMLAHGHNCLDVECLNNGGCQCHTCVNMVMSYECCCK 120
 QY 156 EGFFLSDNQHTCTHRSEGLSCVNMKHGCSHIKEAPRGSVACECPGFELAKNQDCL 215
 DB 121 EGFFLSDNQHTCTHRSEGLSCVNMKHGCSHIKEAPRGSVACECPGFELAKNQDCL 180
 QY 216 TCNHGNGSCGSCDCTADGPECSCHPQYKMHDTGSRCLEREDTVLEVTSTNTTSVVDGDK 275
 DB 181 TCNHGNGSCGSCDCTADGPECSCHPQYKMHDTGSRCLEREDTVLEVTSTNTTSVVDGDK 240
 QY 276 RYKRRLLMETCAVNGGCDRTCKDSTGVHSCSPVGFLLQDGTCKDIDECOTRNGGCD 335
 DB 241 RYKRRLLMETCAVNGGCDRTCKDSTGVHSCSPVGFLLQDGTCKDIDECOTRNGGCD 300
 QY 336 HFCXNVGSEFCGCKGFKLLTDEKSCQVDDESLDRTQDHSCHINHPGTPACACNRYTL 395
 DB 301 HFCXNVGSEFCGCKGFKLLTDEKSCQVDDESLDRTQDHSCHINHPGTPACACNRYTL 360
 QY 396 YGPTHGDTRECSINNGGCGQCVNTVSGYECCHPOYKJLHWNKDCVEYKGLPTSVSP 455
 DB 361 YGPTHGDTRECSINNGGCGQCVNTVSGYECCHPOYKJLHWNKDCVEYKGLPTSVSP 420
 QY 456 RVSUHGCKSGGCGCFLRCHSG:HLSSDVTITRTSVTFKLNKCKSLKNAELFPEGLRPA 515
 DB 421 RVSUHGCKSGGCGCFLRCHSG:HLSSDVTITRTSVTFKLNKCKSLKNAELFPEGLRPA 480
 QY 516 LPEKSSSVKESFRYVNLTCSSGKQVPCAPQRPSTPKEMFITVEFELETNKEVTASCLLS 575
 DB 481 LPEKSSSVKESFRYVNLTCSSGKQVPCAPQRPSTPKEMFITVEFELETNKEVTASCLLS 540
 QY 576 CIVKRTKRLRKA:RTLRKAVHRECEFLQSGMNLQVAKKPPRESERCAESCGVGQGHAE 635
 DB 541 CIVKRTKRLRKA:RTLRKAVHRECEFLQSGMNLQVAKKPPRESERCAESCGVGQGHAE 600
 QY 636 NQVSCRAGTYVDGARERC:LCPNGTFQNEEGQMTCEPCPRPNSGALKTPEANMSECG 695
 DB 601 NQVSCRAGTYVDGARERC:LCPNGTFQNEEGQMTCEPCPRPNSGALKTPEANMSECG 660
 QY 696 GLCOFGYSADGAFACOLCALGTFQPEAGRTSCFPGGGGLATKHGATSPDCETRVQCS 755
 DB 661 GLCOFGYSADGAFACOLCALGTFQPEAGRTSCFPGGGGLATKHGATSPDCETRVQCS 720
 QY 756 PGHFNITTHRCIRCPVGTQYQPFQKNCVSCFQNTITDPDGSNTITQCKNRRCGGELGD 815
 DB 721 PGHFNITTHRCIRCPVGTQYQPFQKNCVSCFQNTITDPDGSNTITQCKNRRCGGELGD 780
 QY 816 PTGVIESPNYPGNYPANTECTWTINPPPKRILIVVPEIFLPEDDGGDYLVMKTSSEN 875
 DB 781 PTGVIESPNYPGNYPANTECTWTINPPPKRILIVVPEIFLPEDDGGDYLVMKTSSEN 840
 QY 876 SVTTYETCCQYVERIAETSESKLWIOFKSNGNSARGPOVYVYDDEYOEELIETVRD 935
 DB 841 SVTTYETCCQYVERIAETSESKLWIOFKSNGNSARGPOVYVYDDEYOEELIETVRD 900
 QY 936 GRLYASENHQEI:KDKKL:KALFDVLAHPQNYFYKTAQESREMPRPSFIRLLRSKVSRL 995
 DB 901 GRLYASENHQEI:KDKKL:KALFDVLAHPQNYFYKTAQESREMPRPSFIRLLRSKVSRL 960
 QY 996 RPYK 999
 DB 961 RPYK 964

RESULT 3
 RAG16642
 ID AAO:6642 standard; Protein; 919 AA.
 XX
 AC AAO16642;
 XX
 DT 10-MAY-2003 (first entry)
 XX Human extracellular messenger protein #2.
 XX

Human; gene therapy; protein replacement therapy;
 extracellular messenger; EXMES; cardiovascular disorder; cancer;
 congestive heart failure; autoimmune disorder; inflammatory disorder;
 AIDS; neurological disorder; epilepsy; reproductive disorder;
 infertility; infection; developmental disorder; Cushing's syndrome;
 endocrine disorder; diabetes insipidus; cell proliferative disorder.
 XX
 OS Homo sapiens.
 XX W02003002610-A1.
 XX 09-JAN-2003.
 XX 26-JUN-2002; 2002WO-US20430.
 XX 29-JUN-2001; 2001US-301789P.
 PR 21-SEP-2001; 2001US-324149P.
 PR 05-OCT-2001; 2001US-327113P.
 PR 12-OCT-2001; 2001US-329215P.
 PR 14-DEC-2001; 2001US-340218P.
 PR 05-APR-2002; 2002US-370761P.
 PR 19-APR-2002; 2002US-373824P.
 XX
 PA (INCY-) INCYTE GENOMICS INC.
 XX
 PI Deleagean AM, Borowsky ML, Khan FA, Kearney L, Ramkumar J;
 PI Malia NK, Lu Y, Honchell CD, Kallick DA, Emerling BM, Gervad AB;
 PI Griffin JA, Warren BA, Yue H, Thangavelu K, Sprague KW, Ison CH;
 PI Elliott VS, Mason PM, Richardson TW, Tran UK, Swarnakar A, Jin P;
 PI Kable AE;
 XX
 DR WPI; 2003-210247/20.
 DR N-PSDB; AAL51944.
 XX
 PT New human extracellular messengers (EXMES) and polynucleotides, useful
 for diagnosing, treating or preventing e.g. congestive heart failure, or
 AIDS, stroke, Alzheimer's disease, infertility, infections, leukemia or
 breast cancer.
 XX
 PS Claim 1; Page 182-184; 243pp; English.
 XX
 CC The invention comprises the amino acid and coding sequences of human
 extracellular messengers (EXMES). The EXMES DNA and protein sequences are
 useful for treating a disease or condition associated with decreased/
 increased expression of functional EXMES. The EXMES DNA and protein
 sequences are useful for diagnosing, treating and preventing:
 CC cardiovascular disorders (e.g. congestive heart failure); autoimmune/
 CC inflammatory disorders (e.g. AIDS); neurological disorders (e.g.
 CC epilepsy); reproductive disorders (e.g. infertility); infections;
 CC developmental disorders (e.g. Cushing's syndrome); endocrine disorders
 CC (e.g. diabetes insipidus); and cell proliferative disorders (e.g.
 CC cancer). The present amino acid sequence represents a human EXMES protein
 CC of the invention.
 XX
 SQ Sequence 319 AA;

Query Match
 Best Local Similarity 90.6%; Score 5079; DB 24; Length 919;
 Marches 916; Conservative 1; Mismatches 2; Indels 80; Gaps 1;

QY 1 MGAGNRNPGAAWAVLLLLPPLLLLAGAVPGRGAGPQEDVDECAQGLDCHADA 60
 DB 1 MGAGNRNPGAAWAVLLLLPPLLLLAGAVPGRGAGPQEDVDECAQGLDCHADA 60
 QY 61 LCQNTPTSYKCSCKPGYQGGROCEDIDECGNELNGGCVHDCNIPGNVRCCTCFDGFMLA 120
 DB 61 LCQNTPTSYKCSCKPGYQGGROCEDIDECGNELNGGCVHDCNIPGNVRCCTCFDGFMLA 120
 QY 121 HDGNCLDVBDECELENNGGCQHTCVNMGSYECCKEGFFLSDNQHTCIHRSEEGLSCKMK 180
 DB 121 HDGNCLDVBDECELENNGGCQHTCVNMGSYECCKEGFFLSDNQHTCIHRSEEGLSCKMK 180
 QY 181 DHGSHLCKEAPRGSVACRPGFELAKNQDCILTCNHNGGCGQSCDDTAGPCECSCH 240

DB 181 DHGSHLCKEAPRGSVACRPGFELAKNQDCILTCNHNGGCGQSCDDTAGPCECSCH 240
 QY 241 PQYKHTDGRSCLEREJTVLEVTESNTTSVVDGDKRVKRRLLMETCAVNNGGCRCTCKDT 300
 DB 241 PQYKHTDGRSCLEREJTVLEVTESNTTSVVDGDKRVKRRLLMETCAVNNGGCRCTCKDT 300
 QY 301 STGVHSCPVGFTLQLDGKTKOIDEQTRNGGCDHFCKNIVGSCGCKKGFLLTDEK 360
 DB 301 STGVHSCPVGFTLQLDGKTKOIDEQTRNGGCDHFCKNIVGSCGCKKGFLLTDEK 360
 QY 361 SCQDVDCSLDRCTCHSCINHPGTFCACNRGYTLGFTHCG----- 402
 DB 361 SCQDVDCSLDRCTCHSCINHPGTFCACNRGYTLGFTHCG----- 402
 QY 421 TVGSYEQCHPGYKLVHWNKKDCVEVKGLLPVTSVSPVSLHCGKSGGGCGGCFLRCHSGIHL 480
 DB 403 ----- 422
 QY 481 SSDVTITRTSVTFKLNKGKSLKNAELFPEGLRPALPEKHSSVKESPRYVNLTCSSGQV 540
 DB 403 --DVTITRTSVTFKLNKGKSLKNAELFPEGLRPALPEKHSSVKESPRYVNLTCSSGQV 460
 QY 541 PGAGRPSTPKEMFITYVEFELETNQKEVTASCDLSCLVKTETELKAIETLEKAVHREQ 600
 DB 461 PGAGRPSTPKEMFITYVEFELETNQKEVTASCDLSCLVKTETELKAIETLEKAVHREQ 520
 QY 601 FHLQSLGNLDVAKKPPRTSERQAESCGVGGHAENCCVSCRAGTYVDGARERCILCPNG 660
 DB 521 FHLQSLGNLDVAKKPPRTSERQAESCGVGGHAENCCVSCRAGTYVDGARERCILCPNG 580
 QY 661 TFQNEEGOMTCEPCPRPNSGALKTPAAMWSECGGLCQPGEYSADGFPACQLCALGTFQ 720
 DB 581 TFQNEEGOMTCEPCPRPNSGALKTPAAMWSECGGLCQPGEYSADGFPACQLCALGTFQ 640
 QY 721 PEAGRTSCFPCCGGGLATKHQATSFQDCETRVQCSGPHFYNTTTHRCIRCPVGTYPQEPFG 780
 DB 641 PEAGRTSCFPCCGGGLATKHQATSFQDCETRVQCSGPHFYNTTTHRCIRCPVGTYPQEPFG 700
 QY 781 KNCVSCPGNTTDFDGSNTITQCKNRRCGELGDFGTGYIESPNYPNPANTECTWTIN 840
 DB 701 KNCVSCPGNSTTDFDGSNTITQCKNRRCGELGDFGTGYIESPNYPNPANTECTWTIN 760
 QY 841 PPPKRRILVWPEFLPIEDDCGYLVNRKTSNNSVTVTETCTYERPIAFTSRSKKLW 900
 DB 761 PPPKRRILVWPEFLPIEDDCGYLVNRKTSNNSVTVTETCTYERPIAFTSRSKKLW 820
 QY 901 IOFKSNEGSNARGFCVPVTVDEDYQELIEDVDRGLYASENHQELKOKKLKALFDV 960
 DB 821 IOFKSNEGSNARGFCVPVTVDEDYQELIEDVDRGLYASENHQELKOKKLKALFDV 880
 QY 961 LAHPQNYFKYTAQESREMFPSFIRLLSKVSRELRPYK 999
 DB 881 LAHPQNYFKYTAQESREMFPSFIRLLSKVSRELRPYK 919

RESJUT IC
 ABBE927
 ID ABBE927 standard; Protein; 597 AA.
 XX
 AC ABBE927;
 XX
 DT 06-OCT-2002 (first entry)
 XX
 DE Mouse BCO2 orthologue protein.
 XX
 KW Breast cancer modulating protein; BCMP; BCO2; cytostatic; mouse;
 KW immunostimulant; antisense therapy; gene therapy; vaccine.
 XX
 OS Mus sp.
 XX
 EN WC0200255988-A2.

PD 18-JUL-2002.
 XX
 PF 21-DEC-2001; 2001WO-US05048.
 XX
 PR 21-DEC-2000; 2000US-0747371.
 XX (EOSB-) EOS BIOTECHNOLOGY INC.
 XX
 PI Gish KC, Mack D;
 XX
 DR WPI; 2002-566749/60.
 XX
 PT Diagnosing breast cancer comprises comparing the gene expression of a
 PT breast cancer modulating protein in a test breast tissue and unaffected
 PT breast tissue of the same or a different patient -
 XX
 PS Disclosure; Fig 4A-B; 68pp; English.
 XX
 CC The invention relates to diagnosing breast cancer that involves
 CC determining the expression of a gene encoding breast cancer modulating
 CC protein (BCO2) or a fragment of it in a breast tissue of an individual,
 CC and comparing the expression of the genes from a second normal breast
 CC tissue from the individual or a second unaffected individual, where a
 CC difference in expression indicates the individual has breast cancer. The
 CC method is used for diagnosing breast cancer. Other new methods are
 CC provided for screening drug candidates, screening for bioactive agents,
 CC and evaluating the effect of a candidate breast cancer drug. An antibody
 CC to BCO2 can be used to inhibit the bioactivity of BCO2 or neutralise it's
 CC effect and can be used to treat breast cancer. The antibody can also be
 CC used to localise a therapeutic group to breast cancer tissue, which can
 CC be used to treat breast cancer. Antisense molecules are used to inhibit
 CC breast cancer in a cell. BCO2 protein and polynucleotides are used to
 CC elicit an immune response in an individual. BCO2 is also used to
 CC determining the prognosis of an individual with breast cancer. The
 CC present sequence represents the mouse BCO2 orthologue protein.
 XX
 SQ Sequence 997 AA;

Query Match 90.2%; Score 5056; DB 23; Length 997;
 Best Local Similarity 89.2%; Pred. No. 8e-306;
 Matches 891; Conservative 47; Mismatches 59; Indels 2; Gaps 1;

QY 1 MGVAQRNRGAAMVLLLLLLPPLLLLAGAVPPGRGAAGPQEDVDECAQGLDCHADA 60
 DB 1 XGVACGRPREARALLLLLLLPP--LLAAVPPDRGUTNGPSEVDVDECAQGLDCHADA 58
 QY 61 LCQNTPTSYKCSKPGYQEGRCQEDIDECGNELNGGCVHDCNLINPGNYRCTCFDGFMLA 120
 DB 59 LCQNTPTSYKCSKPGYQEGRCQEDIDECGNELNGGCVHDCNLINPGNYRCTCFDGFMLA 118
 QY 121 HDGHNCLDVECLNNGGCGHCTVNVMSGYECCKEGFFLSNQHCTCHRSSEGLSCMNK 180
 DB 119 HDGHNCLDVECLNNGGCGHCTVNVMSGYECCKEGFFLSNQHCTCHRSSEGLSCMNK 178
 QY 181 DHGCSHICKEAPRGSVACRCRPGFELAKNQRCDILCNHNGGCGHSCDDTADGPECSC 240
 DB 179 DHGCGHICKEAPRGSVACRCRPGFELAKNQRCDILCNHNGGCGHSCDDTADGPECSC 238
 QY 241 PQYKMTDGRSLEREDTVLEVTESNTTVSDGDKRVKRLLMETCAVNVGSCDRCTCKDT 300
 DB 239 PRYRLHADGRSLCEQGTVEGTESNATSVADGDKRVKRLLMETCAVNVGSCDRCTCKDT 298
 QY 301 STGVHSCVPVGTLOJLDGKTCKDIDECQTRNGGCHFCNKIVGSEDCGCKGFKLLIDEX 360
 DB 299 STGVHSCVPVGTLOJLDGKTCKDIDECQTRNGGCHFCNKIVGSEDCGCKGFKLLIDEX 358
 QY 361 SCQVDDECSLDRTCDHSCINHPGTACACNRGYTLVGFTHCGDTNECSINNGCGQOVN 420
 DB 359 SCQVDDECSLDRTCDHSCINHPGTACACNRGYTLVGFTHCGDTNECSINNGCGQOVN 418
 QY 421 TVGSYECQCHPGYKHLWKKDCVVEKGLLPTSVPRVSJHCCKSGGGGCGCFURCHSGH 480
 DB 419 TVGSYECQCHPGYKHLWKKDCVVEKGLLPTSVPRVSJHCCKSGGGGCGCFURCHSGH 478

QY 431 SSDVTTIRTSVTEKLNKSGKSLKNAELFPEGLRPALPEKHSSVSKESFRYVNLTCSSQKQV 540
 DB 479 SSDVTVRTSVTEKLNKSGKSLKNAELFPEGLRPALPEKHSSVSKESFRYVNLTCSSQKQV 538
 QY 541 PGAPGRPSTPEMFITVEFELETNQKEVTASCLSCIVKRTKRLKRAIRTLKRAHREQ 600
 DB 539 PGALGRINAPKEMFITVEFELETNQKEVTASCLSCIVKRTKRLKRAIRTLKRAHREQ 598
 QY 601 FHLQSGMNLJVAKKPPTISERCAESCGVGQGHAEHQVSCRASTYDQARERCILCPNG 660
 DB 599 FHLQSGMNLJVAKKPPTISERCAESCGVGQGHAEHQVSCRASTYDQARERCILCPNG 659
 QY 661 TFCNEGQMTCEPCPRFGNSGALKTPRANMNSGGGLGCPGEYSADGAPCQJCALGTFO 720
 DB 659 TFCNEGQMTCEPCPRFGNSGALKTPRANMNSGGGLGCPGEYSADGAPCQJCALGTFO 718
 QY 721 PEAGRTSCPPCGGLATKHQATSFQDCETRVCCSPGRFYNTHRCIRCPVGTVOPEFG 780
 DB 719 PDVGRSCLSCGGGLPTKHLGATSFQDCETRVCCSPGRFYNTHRCIRCPVGTVOPEFG 778
 QY 781 KKNVSCPGNTTTFDQSTNITQCKNRRCGGELGDTGYIESPNYPNPANTECTWTIN 840
 DB 779 KKNVSCPGNTTTFDQSTNITQCKNRRCGGELGDTGYIESPNYPNPANTECTWTIN 838
 QY 841 PPKERLLIVPEIFPIEDDCCGDLVNRKTSNSNVTTVCTCTYERP-AFTSRKKLW 900
 DB 839 PPKERLLIVPEIFPIEDDCCGDLVNRKTSNSNVTTVCTCTYERP-AFTSRKKLW 898
 QY 901 IQFKSNEGNSARGFQVYVYTYDEYQGLJEDIVRDGRLYASENHQELKDKKLIKALFDV 960
 DB 899 IQFKSNEGNSARGFQVYVYTYDEYQGLJEDIVRDGRLYASENHQELKDKKLIKALFDV 958
 QY 961 LAHPQNYFKYTAQESREMFPRSPFIRLLRSKVSRLRPYK 999
 DB 959 LAHPQNYFKYTAQESREMFPRSPFIRLLRSKVSRLRPYK 997

RESULT 11
 ABP56758
 ID ABP56758 standard; Protein; 957 AA.
 XX AC ABP56758;
 XX DT 31-XAR-2003 (first entry;
 XX DE Mouse CEGP1 protein SEQ ID NO:6.
 XX KW Human; secreted protein; epidermal growth factor protein subfamily;
 KW epidermal growth factor; EGF; Gene therapy; therapeutic; drug screening;
 KW biological activity; immune response.
 XX CS Mus musculus.
 XX FN W0200210:080-A2.
 XX PD 19-DEC-2002.
 XX PF 07-MAY-2002; 2002WC-US22278.
 XX PR 16-MAY-2001; 2001US-0855824.
 XX PA (PEKE) BE CORP NY.
 PA (GONG); GONG F.
 PA (CFRA); DI FRANCESCO V.
 XX PI Gong F, Di Francesco V, Van C, Beasley EM;
 XX WPI; 2003-167407/16.
 XX
 PT New isolated human secreted peptides, useful for diagnosing or treating
 PT a disease characterized by an absence of, inappropriate or unwanted
 PT expression of the secreted protein, and in drug screening assays -

XX	PS	Disclosure: Fig 2f-G; 185pp; English.	
XX	CC	The present sequence encodes a human secreted peptide (I) related to the epidermal growth factor (EGF) protein subfamily. (I) can be used in gene therapy. (I) and the nucleic acid molecules encoding (I) can be used as models for the development of human therapeutic targets, to aid in the identification of therapeutic proteins, and to serve as targets for the development of human therapeutic agents. (I) may be used in drug screening assays, in assays to determine the biological activity of the protein, to raise antibodies or to elicit another immune response, as a reagent in assays designed to quantitatively determine levels of the protein in biological fluids, or as markers for tissues in which the corresponding protein is preferentially expressed. (I) can also be used for diagnosing or treating a disease or disorder characterized by an absence of, inappropriate or unwanted expression of the protein. (I) is located to human chromosome 22. The present sequence represents mouse CEGP1 protein which is given in comparison with (I) in the exemplification of the present invention.	
XX	XX	Query Match 88.0%; Score 4935; DB 24; Length 957; Best Local Similarity 90.2%; Pred No. 2.5e-298; Matches 862; Conservative 46; Mismatches 48; Indels 0; Gaps 0;	
QY	44	EDVDECAQGLDCHADALCONFTSYKSCKPGYQGEGRQCEIDDECQNELNGGCVHDC	103
Db	2	EDVDECAQGLDCHADALCONFTSYKSCKPGYQGEGRQCEIDDECQNELNGGCVHDC	61
QY	104	NIPGNRYCTCFGFMHLADHGNCLDVECLNNGGCGHTCVNNGSYECCKEFGFLSDN	163
Db	62	NIPGNRYCTCFGFMHLADHGNCLDVECLNNGGCGHTCVNNGSYECCKEFGFLSDN	121
QY	164	QHTCHIRSEGLSCNKHGCHICKERAPRGSVACCEPGFELAKNOROCILTCNHNGG	223
Db	122	QHTCHIRSEGLSCNKHGCHICKERAPRGSVACCEPGFELAKNOROCILTCNHNGG	181
QY	224	QHSQCDTADGPECSCHPOYKMTDORSCLEREDTVLEVTESNTTSVVDGKVRKRLLM	283
Db	182	QHSQCDTADGPECSCHPYRLHADGRSCLEQEGTVLEGTESNATSVADGKVRKRLLM	241
QY	284	ETCAVANGGCDRTCKTSTGVHCSCPGVFTLQDKTKDIDECQTRNGGCDHFCNIVG	343
Db	242	ETCAVANGGCDRTCKTSTGVHCSCPTGFTLQVCGKTKDIDECQTRNGGCHFCNIVG	301
QY	344	SFDCGCKGFKLLTDEKSCQDVDECSLDRTCDSHGINHPGTFCACACNRYGLYGFTHCGD	403
Db	302	SFDCSCKGFKLLTDEKSCQDVDECSLDRTCDSHGINHPGTFCACNPGYLYSTHCGD	361
QY	404	TNECSINNGGCGQCVNTVGSYECQCHPGYKLFHNKDCVVKGLPTSVSPRSLHCGK	463
Db	362	TNECSVNGGCGQCVNTVGSYECQCHPGYKLFHNKDCVVKGLPTSVSPRSLHCGK	421
QY	464	SGGDCGFLRCHSGHLSDVTTIRTSVTFKLENGKSLKNAELFPEGRLPALPEHSSV	523
Db	422	SGGDRCFRLRCHSGHLSDVTTIRTSVTFKLENGKSLKNAELFPEGRLPALPEHSSV	481
QY	524	KESFVNLTCSSGKQVPCAPGRSPTEKMFITVEFELETNCKEVTASCDLSICVKTREK	583
Db	482	KESFQVNLTCSPGKQVPCALGRNALPKEMFITVEFERTEYKEVTASCNLSVVKREK	541
QY	584	RLKALIRTKVAHREQFLQLSGMNLDAKPPPTSERQASCGVGQGHAEINQCVSRA	643
Db	542	RLKALIRTKVAHREQFLQLSGMDLMAKTPSRVSGQHEETCGVGQGHESQCVSRA	601
QY	644	GTYDGAERCLICPNTQNEGQWTCPCPRPGNSGALKTPPEANNSECCGLCPQGEY	703
Db	602	GTYDGSQERCLICPNTQNEGQWTCPCPRPENLGSJKISEANNVSDCGGLCPQGEY	661
QY	704	SADGFAPCALGTFTQPEAGRTSCFPCCGGGLATKHGATSFQDCETRVCQSPGHFYNTT	763
Db	662	SANGFAPCALGTFTQPDVGRITSCLSGCGGLPTKHLGATSFQDCETRVCQSPGHFYNTT	721

QY	764	THRCIRCPVGTGPEFSKNCVSCPGNTTDFDGSNTTCKNRRCGGELGDTFTGYESP	823
Db	722	THRCIRCPVGTGPEFSKNCVSCPGNTTDFDGSNTTCKNRRCGGELGDTFTGYESP	781
QY	824	NYGKNYPANTECTWTINPPKRRILLVPEIPLIEDCGDYVMRKTSSSNSVTYETC	883
Db	782	NYGKNYPANSECTWTINPPKRRILLVPEIPLIEDCGDYVMRKTSSSNSVTYETC	841
QY	884	QYVERIAFTSRSKKWIQFKSNEGNSARGFOVYVYDCEYCELIEDVRGRLYASEN	943
Db	842	QYVERIAFTSRSKKWIQFKSNEGNSARGFOVYVYDCEYCELIEDVRGRLYASEN	901
QY	944	HQBIKDKKLKALFDVLAHPQNYFYTACEREMPPRPFIRLLSKVSRFLRPYK	999
Db	902	HQBIKDKKLKALFDVLAHPQNYFYTACEREMPPRPFIRLLSKVSRFLRPYK	957
XX	RESULT 12		
XX	ID	ABP56755 standard; Protein; 988 AA.	
XX	AC	ABP56755;	
XX	DT	31-MAR-2003 (first entry)	
XX	XX	Human epidermal growth factor related secreted protein SEQ ID NO:2.	
XX	XX	Human; secreted protein; epidermal growth factor protein subfamily;	
XX	XX	epidermal growth factor; EGF; gene therapy; therapeutic; drug screening;	
XX	XX	biological activity; immune response; chromosome 22.	
XX	XX	Homo sapiens.	
XX	XX	MO200210:080-A2.	
XX	XX	19-DEC-2002.	
XX	XX	07-MAY-2002; 2002WD-US22275.	
XX	XX	16-MAY-2001; 2001US-0855624.	
XX	XX	(PEKE) PE CORP NY.	
XX	XX	(GONG) GONG F.	
XX	XX	(DPPA) D: FRANCESCO V.	
XX	XX	Gong F, Di Francesco V, Yan C, Beasley EM;	
XX	XX	WPI; 2003-167407/16.	
XX	XX	N-PSDB; AB222653, AB222654.	
XX	XX	New isolated human secreted peptides, useful for diagnosing or treating a disease characterized by an absence of, inappropriate or unwanted expression of the secreted protein, and in drug screening assays -	
XX	XX	Claim 1; Fig 2A; 185pp; English.	
XX	XX	The present sequence represents a human secreted peptide (I) related to the epidermal growth factor (EGF) protein subfamily. (I) can be used in gene therapy. (I) and the nucleic acid molecules encoding (I) can be used as models for the development of human therapeutic targets, to aid in the identification of therapeutic proteins, and to serve as targets for the development of human therapeutic agents. (I) may be used in drug screening assays, in assays to determine the biological activity of the protein, to raise antibodies or to elicit another immune response, as a reagent in assays designed to quantitatively determine levels of the protein in biological fluids, or as markers for tissues in which the corresponding protein is preferentially expressed. (I) can also be used for diagnosing or treating a disease or disorder characterized by an absence of, inappropriate or unwanted expression of the protein. (I) is located to human chromosome 22.	
XX	XX	Sequence 988 AA;	

Qy	10	GAA---WAVLLLLPPLLLLAGAVPPGRRAAG-----POEDVDECAGLDDCHADAL	62
Db	2	GAAAVRHLCVLLAL-----GTRGLAGSGUGPSVDDESGTDDCHDAI	49
Qy	62	CONTPTSYKSCKPGYQGEGRQCEIDECGNE-LNGGCVDFCLNIPGNVRCCTCFDGFMLA	120
Db	50	QNTPKSYKCLKRPGYKGBGKQCEIDECENDYNGCVHCEINIPGNVRCCTCFDGFMLA	129
Qy	121	HDGNCLDVDECEJNNGGCOHTCVNVMGSEYECCKEGFFLSNQHCTCHRSEGLSOMNK	180
Db	110	HDGNCLDVDECEJNNGGCOHTCVNVMGSEYECCKEGFFLSNQHCTCHRSEGLSOMNK	169
Qy	181	CHGSHICKAPRGSVACRCPGFELAKNORCILTCHNNGGCGHSCDDTADGPECSC	240
Db	170	DHCAHICRETPKGVACDCRPGFDLAQOKOCTLTCHNNGGCGHSCDDTADGPECSC	229
Qy	241	PQYKMTDGRSCLEREDTVLEVTESNTTSVWDGKVKRRLXMETCAVNNGCCDRTCKDT	300
Db	230	QKVALSHSGRTCI-----ETCAVNNGCCDRTCKDT	259
Qy	301	STGVHSCSPVGTFLDLDGKTKCDIDECQTRNGCOHCFKNI VGSFDCCKKGFKLCDEK	360
Db	260	ATGVRCSPVGTFLDLDGKTKCDIDECQTRNGCOHCFKNTVGSFDCCKKGFKLCDEK	319
Qy	361	SCQDVDECSLDRTCHSCINHPGTFACACNRGYTLVGFTHGCTNECSINNGCQGVGVN	420
Db	320	TCQDIDECSPERTCHDHCINSPGSCFCLCHRGVILYGTTHCGDVBDECSNMGSCDGGCVN	379
Qy	421	TVGSYECQCHPGYKLNKNDKVEV-KGILLPTSVSPRVSHCGKSGGDCGFLRC-----	474
Db	380	TGSEYECVCPGRRLLHWKDCVETGKLSRAKTSFRAQLSCSKAGGVESCFSPANTL	439
Qy	475	-----HSGTHLS---SDVTITRTSVTFKLNKGKSL	502
Db	440	FVPDSSENSVYLCGVPGPOGKALQKNGTSSGLSPSCSDAPTPIKQKARFKIRDAKCHL	499
Qy	503	KNAELFPEGRLPALPKHSSVKESFR-----YVNLTCSSQKQVPGAPGRSTPK	552
Db	500	R-----PHSQARAKETARQPLDHOHVFTVTKCDSSKK--RRGRKSPSK	544
Qy	553	M-FITVEFEITNQEVTASCDLSCIVKRTKELRKAIITLRKAVHREOFHLQSGMNL	611
Db	545	VSHITAEFEITKWEASDTCADCLRKRAEOSLOAAIKLRS-GRQFYVVGSGTEY	604
Qy	612	VAKKPTSRQAESCQVGGQHAENQCVSCRAGTYVYDGBARECILCPNGTPONEGQVTC	671
Db	605	VAORPAKALEGGQ-ACAGAGVLQDSKVCACGPGTHFGBELGCVSCMPTGYDMEGQLSC	663
Qy	672	EPCPRGNSGALKTPAAMNMBEGGLCQCPGEYSADGFAPCQCALGTFQPEAGRTSCFPC	731
Db	664	TPCP---SSDGLGLPGARNVSECGGQCSGFFSADGFKPCQACPVYQPEGRGCFPC	720
Qy	732	GGGLATHQAGATFQDCETVQCSQGHVYNTTHRCIRCPVQTYQPEFGKNNVSCPKNT	791
Db	721	GGGLTKHEGTTSFQDCEAKVHCSPGHHVNTTHRCIRCPVQTYQPEFGKNNVSCPKNT	780
Qy	792	TTDFGSGTNITCKNRCGGELEDFTGYTESPNVPGNYPANFECTWTNFPKRLIVV	851
Db	781	STDFGSGTNVTHCKNQKGGELGDTGYTESPNVPGNYPANFECTWTNFPKRLIVV	840
Qy	852	PEIFLIEDCGDYLVMKRTSSNSVTTTETCTYERPIAFTSRKSLMIQPKSNEGNSA	911
Db	841	PEIFLIEDCGDYLVMKRSASPTSIITVETQCTYERPAFTSRKSLMIQPKSNEGNSG	900
Qy	912	RGQVPPVTVDDYQELIEDIVDRGLYASENHQELKDKKLIKALFDVLAPHPQNYFYK	971
Db	901	KGQVPPVTVDDYQYQIEDIVDRGLYASENHQELKDKKLIKALFDVLAPHPQNYFYK	960

Qy	972	AQESREMPFRSPFIRLLRSKVSRLRPYK	999
Db	961	AQESREMPFRSPFIRLLRSKVSRLRPYK	988
RESULT 13			
AB004645			
CD	AB004649	standard; Protein; 988 AA.	
XX	ABJ04649;		
AC	ABJ04649;		
DT	11-OCT-2002	(first entry)	
XX	Protein of NOVX ab SEQ ID No 16.		
DE			
XX	Cytostatic; antidiabetic; anorectic; anorectic; metabolic; rootropic; antilipemic;		
KW	neuroprotective; antiparkinsonian; anticonvulsant; cerebroprotective;		
KW	transluciser; neuroleptic; antidiabetic; antitumor; antitumor; antitumor;		
KW	anti-HIV; antiallergic; antirheumatic; antitumor; antitumor; antitumor;		
KW	metabolic disorder; obesity; infectious disease; Alzheimer's disease;		
KW	anorexia; neurodegenerative disorder; Parkinson's disorder; obesity;		
KW	immune disorder; haematopoietic disorder; dyslipidaemia; chronic disease;		
KW	metabolic syndrome X; wasting disorder; cancer; neurological disorder;		
KW	epilepsy; stroke; mental disorder; schizophrenic disorders; goiter;		
KW	vesicular transport; cystic fibrosis; gastrointestinal disorder;		
KW	diabetes mellitus; ulcerative colitis; AIDS; allergic reaction;		
KW	multiple sclerosis; rheumatoid arthritis; transgenic animal;		
XX	gene therapy.		
XX	Unidentified.		
OS			
XX	W0200246409-A2.		
PN			
XX	13-JUN-2002.		
PD			
XX	06-DEC-2001; 2001WO-US46586.		
PF			
XX			
PR	06-DEC-2000; 2000US-251660P.		
PR	12-DEC-2000; 2000US-255029P.		
PR	08-JAN-2001; 2001US-260326P.		
PR	24-JAN-2001; 2001US-263800P.		
PR	20-FEB-2001; 2001US-269942P.		
PR	24-APR-2001; 2001US-286183P.		
PR	23-AUG-2001; 2001US-313627P.		
PR	12-SEP-2001; 2001US-318712P.		
XX	(CURA-)	CURAGEN CORP.	
XX			
XX	Guo X, Li L, Patturajan M, Shimkets RA, Casman SJ, Malyankar UM;		
PI	Chernov V, Vernet CA, Spytek KA, Shenoy SG, Alsbrook JF;		
P	Edinger S, Peyman CA, Stone DC, Eilerman K, Gargoli EA;		
P	Boldog P, Colman SD, Eisen AJ, Liu X, Padigaru M, Spaderna SK;		
PI	Zerhusen BD;		
XX			
XX	WPI: 2002-347774/58.		
DR	N-PSDB; A3705461.		
XX			
PT	Novel isolated polypeptide, designated NOVX, useful for treating or		
PT	preventing cancer, diabetes, obesity, dyslipidemia, anorexia, and		
PT	metabolic, neurodegenerative, immune and hematopoietic disorders		
XX			
PS	Claim 1; Page 69; 421pp; English.		
XX			
CC	The invention relates to an isolated polypeptide, designated NOVX,		
CC	comprising a sequence fully defined in the specification. The isolated		
CC	protein, its encoding polynucleotide or an antibody created from the		
CC	protein is useful in the manufacture of a medicament for treating a		
CC	syndrome associated with a human disease, preferably a NOVX-associated		
CC	disorder, or for treating or preventing a NOVX-associated disorder in a		
CC	subject, preferably human. The isolated protein, its encoding		
CC	polynucleotide or an antibody created from the protein are also useful		
CC	for treating or preventing metabolic disorders, diabetes, obesity,		
CC	infectious disease, anorexia, neurodegenerative disorder, Alzheimer's		

CC disease, Parkinson's disorder, immune disorders, haematopoietic
CC disorders, and various dyslipidaemias, metabolic disturbances associated
CC with obesity, the metabolic syndrome X, wasting disorders associated with
CC chronic diseases, and cancer. The isolated protein, its encoding
CC polynucleotide or an antibody created from the protein are useful for
CC treating or preventing neurological disorders such as epilepsy, stroke,
CC mental disorders including mood, anxiety, schizophrenic disorders,
CC disorders of vesicular transport such as cystic fibrosis, diabetes
CC mellitus, goiter, gastrointestinal disorders including ulcerative
CC colitis, other conditions associated with abnormal vesicle trafficking
CC including AIDS, allergic reactions, multiple sclerosis and rheumatoid
CC arthritis. A cell comprising the vector of the invention is useful for
CC producing non-human transgenic animals. The polynucleotide of the
CC invention can be used to treat disorders by gene therapy. This sequence
CC represents one of the isolated NOVX proteins of the invention.

XX
SQ Sequence 988 AA;

Query Match 65.3%; Score 3662.5; DB 23; Length 988;
Best Local Similarity 62.2%; Pred. No. 3e-219;
Matches 652; Conservative 113; Mismatches 164; Indels 119; Gaps 14;

QY 10 GAA--WAVL-----PQLLLLAGVPPGCRAG-----PQEDVDECAOGLDCHADAL 61
DB 2 GAAAVRHLVCLAL-----GTRGLAGGLPGSDVDECSGEGDOCHDAI 49

QY 62 QNTPTSYKSCPKGYQEGRQCEIDECGNE--LNGGCVHDCILIPGNYRCTCFDGFMLA 120
DB 110 HDGHNCLDVECDQNNNGCQQTICVAMSGYECQCHSGFF--SDNOHTCIHRSNMGNCMK 169

QY 121 HGHNCLDVECDQNNNGCQQTICVAMSGYECQCHSGFF--SDNOHTCIHRSNMGNCMK 180
DB 110 HDGHNCLDVECDQNNNGCQQTICVAMSGYECQCHSGFF--SDNOHTCIHRSNMGNCMK 169

QY 181 DHGCHSHICKEAPRGVACRPFELAKNQDCILTCNHNGGCGHSCDITADGPECSCH 240
DB 170 DHGCAHICRETPKGVACDRPFELAKNQDCILTCNHNGGCGHSCDITADGPECSCH 229

QY 241 POKMHTDGRSCLEREDTIVLETSNTTSVVDGDKVRKRLMETCAVNGGCDRCCKOT 300
DB 230 QKYAPHSDGRTCI-----ETCAVNGGCDRCCKOT 259

QY 301 STGVHSCSPVGTLLDGLDKTKDIDECOTRNGGDHFCNTVGSFDCGCKGKGLLTDEK 360
DB 260 ATGVRCSPVGTLLDGLDKTKDIDECOTRNGGDHFCNTVGSFDCGCKGKGLLTDEK 319

QY 361 SCQDVDECSLDRDCHSCINHPGTACACNRYTLGFTHCGDTNECS--NNGGCGQCVVN 420
DB 320 TCQDIDEGSFERTCDHICINSPGSPQCLCHRGYLYLGTTHCGDVDECSNNGSCDQCVN 379

QY 421 TVGSYECOCCHPGYKLHNKKDCVEV--KGLLP--SVSPVSLHCKSGGDDGCLPC----- 474
DB 380 TKGSECVCPGRRUHRNKKDCVEV--KGLLP--SVSPVSLHCKSGGDDGCLPC----- 439

QY 475 -----HSGIHLS-----SDVTIRTSVTFKLNKSCSL 502
DB 440 FVPDSENSYVLSCGVPGQKALQKNGTSSGLGSPSCSDAPTTPKCKAREKINDAKCHL 439

QY 503 KNAELFPEGLRPALEKUSVKEFR-----YVNLTCSSGKQVPGAPRSTPKPE 552
DB 500 R-----PHSOARAKETARQPLLDHCHVTFTLKCDSSKK--RRGRKSPSKE 544

QY 553 M-FITVEFELETNQEVATSCDLSICIVARTEKRLKRLKALRT--RKAVHREQFHLQSLGNKD 611
DB 545 VSHI--AEFIEI--KWEASDTCBADCLRKRAQSLQAALTKLXIGIQQYVQVUSGTEYE 604

QY 612 VAKKPPRTSERAECSGVGGAHNEQVSCSAGTYVGGARERCILCPNGTQFNEBGMTC 671
DB 605 VAQRPAAKALEGGQ--ACGAGQVQLQSKVACVPGTHFGGELGQCVPCMPGTQVDMEGQLSC 663

QY 672 EPCPRFGNSGALKTPAAMNSECGLGCPGEYSADGAPCOLCALGTQFQEAERTSCFPC 731
DB 664 TPCP-----SSDGLGLPGARNVSECGGQCGPFGFSADGFKPCACPVGTQVPEPGR--TCFPC 720

QY 732 GGLATKHGATSFQDCETRVQSPGHFYNTTTHRCIRCPVGTYQPEFGKNCVSCPNT 731
DB 721 GGLATKHGATSFQDCETRVQSPGHFYNTTTHRCIRCPVGTYQPEFGKNCVSCPNT 731

QY 792 TTDGGSNTITQCKNRRGCGELGDTGY--ESPNYPCNYPAN--ECTWTINPPKRRILIV 851
DB 781 STDGSGNTVTHCNQCHQCGELGDTGY--ESPNYPCNYPAN--ECTWTINPPKRRILIV 840

QY 852 PEIPLPIEDCGVLYNWKTSNNSVTTVETCTYERPIAF--SRSKKLMQPKSNEGNSA 911
DB 841 PEIPLPIEDCGVLYNWKTSNNSVTTVETCTYERPIAF--SRSKKLMQPKSNEGNSA 900

QY 912 RGFQVPVTVYDQYQELIEDIVDGRLYASEMHCILKDKKLILK--FDVLAKPQNYFKYT 971
DB 901 KGFQVPVTVYDQYQELIEDIVDGRLYASEMHCILKDKKLILK--FDVLAKPQNYFKYT 960

QY 972 AQESKREMPFSPFIRLLSKSVSRFLRPYK 999
DB 961 AQESKREMPFSPFIRLLSKSVSRFLRPYK 988

RESULT 14
AAU79172
CD AAU79172 standard; Protein; 1006 AA.
XX
AC AAU79172;
XX
CT 02-JUL-2002 (first entry)
XX
DE Human MEGF/FIBRILLIN-like protein NOV8.
XX
KW Human; MEGF/FIBRILLIN-like protein; receptor; NOVX; NOV8;
KW vaccine; cancer; tumour; bone disorder; avascular necrosis; allergy;
KW haematopoietic disorder; immune disorder; endometriosis; renal disease;
KW infection; inflammatory disease; lung disease; scleroderma; ataxia;
KW bowel disease; appendicitis; blood disorder; cardiovascular disorder;
KW graft versus host disease; GVHD; lymphoedema; brain disorder;
KW ocular disorder; hepatitis C virus infection; cardiac disorder;
KW autosomal dominant deafness; DFNA-2; chromosome 6.
XX
OS Homo sapiens.
XX
EN NC020214368-A2.
XX
PD 21-FEB-2002.
XX
PF 16-AUG-2001; 2001WC-US25624.
XX
PR 16-AUG-2001; 2000US-225692P.
PR 16-AUG-2001; 2000US-225693P.
PR 16-AUG-2001; 2000US-225837P.
PR 18-AUG-2001; 2000US-226236P.
PR 18-AUG-2001; 2000US-226353P.
PR 22-AUG-2001; 2000US-227385P.
PR 23-AUG-2001; 2000US-227395P.
PR 24-AUG-2001; 2000US-227492P.
PR 24-AUG-2001; 2000US-227600P.
PR 14-XAR-2001; 2001US-275952P.
XX
FA (CURA-) CURAGEN CORP.
XX
XX Zethusen BD, Padigaru M, Spytak KA, Spaderra SK, Gangolli EA;
PI Rastelli Z, Burgess CE, Matumder K, Shirkers R, Mishra V;
PI Vernet CAM, Szekeres ES, Grosse WX, Alcobrook JP, Liu X;
PI Gerlach VB, Ellerman K, Smithson G, Peyman J, Stone D;
PI Macdougall J;
XX
DR WPI; 2002-329571/36.
DR N-PSDB; ABK48392.
XX
XX Novel cytoplasmic, nuclear membrane bound and secreted NOVX
PT polypeptides, useful for treating cancers and tumours, bone disorders,

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 22, 2003, 11:30:24 ; Search time 92 Seconds
(without alignments)
2040.161 Million cell updates/sec

Title: US-09-747-371-2

Perfect score: 5605
Sequence: 1 MGVRNRRFGAAWAVLLLLL.....PRSFIRLLRSKVSRLRPRYK 999

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 629382 seqs, 167460630 residues

Total number of hits satisfying chosen parameters: 629382

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA.*

- 1: /cgn2_6/ptodata/2/pubpaa/us07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NPK_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5605	100.0	999	US-09-747-371-2	Sequence 2, Appli
2	5605	100.0	999	US-10-177-293-45	Sequence 45, Appli
3	5581	99.6	997	US-10-176-847-50	Sequence 50, Appli
4	5428	96.8	964	US-09-855-824-5	Sequence 5, Appli
5	5056	90.2	997	US-09-747-371-3	Sequence 3, Appli
6	4935	88.0	957	US-09-855-824-6	Sequence 6, Appli
7	3675.5	65.6	988	US-09-855-824-2	Sequence 2, Appli
8	3675.5	65.6	988	US-10-195-142-2	Sequence 2, Appli
9	3417.5	61.0	974	US-09-898-570-14	Sequence 14, Appli
10	3417.5	61.0	974	US-09-839-446-14	Sequence 14, Appli
11	3397	60.6	993	US-09-239-663-36	Sequence 36, Appli
12	3394	60.6	1009	US-09-898-570-16	Sequence 16, Appli
13	3394	60.6	1009	US-09-839-446-16	Sequence 16, Appli
14	3290.5	58.7	964	US-10-239-663-58	Sequence 58, Appli
15	3290.5	58.7	965	US-10-221-097-51	Sequence 51, Appli

16	3282.5	58.2	911	12	US-09-855-824-4	Sequence 4, Appli
17	2931.5	52.3	845	10	US-09-898-570-12	Sequence 12, Appli
18	2931.5	52.3	845	11	US-09-839-446-12	Sequence 12, Appli
19	2909	51.9	897	12	US-10-239-663-35	Sequence 35, Appli
20	2819	50.3	785	15	US-10-195-142-8	Sequence 8, Appli
21	2439.5	43.5	735	10	US-09-898-570-10	Sequence 10, Appli
22	2439.5	43.5	735	11	US-09-839-446-10	Sequence 10, Appli
23	2162	38.6	392	10	US-09-898-570-39	Sequence 39, Appli
24	2162	38.6	392	11	US-09-839-446-39	Sequence 39, Appli
25	1680	33.5	594	15	US-10-195-142-10	Sequence 10, Appli
26	1826	32.6	487	12	US-10-221-097-50	Sequence 50, Appli
27	1615.5	28.8	418	10	US-09-886-429-2	Sequence 2, Appli
28	1615.5	28.8	418	15	US-10-162-435-21	Sequence 21, Appli
29	1239.5	22.1	333	15	US-10-195-142-4	Sequence 4, Appli
30	1091	19.5	421	12	US-10-017-161-1606	Sequence 1606, Ap
31	1055.5	18.8	250	15	US-10-195-142-6	Sequence 6, Appli
32	978.5	17.5	286	15	US-10-195-142-12	Sequence 12, Appli
33	945	16.6	387	12	US-10-239-663-57	Sequence 57, Appli
34	932.5	16.6	208	9	US-09-764-853-683	Sequence 683, App
35	932.5	16.6	211	11	US-09-764-881-114	Sequence 114, App
36	927.5	16.5	224	9	US-09-764-853-648	Sequence 648, App
37	927.5	16.5	224	11	US-09-764-881-162	Sequence 162, App
38	681	12.1	776	14	US-10-600-512-8	Sequence 8, Appli
39	681	12.1	959	14	US-10-600-512-10	Sequence 10, Appli
40	635	11.3	915	10	US-09-909-320-34	Sequence 34, Appli
41	635	11.3	915	10	US-09-909-088B-34	Sequence 34, Appli
42	635	11.3	915	10	US-09-905-201A-34	Sequence 34, Appli
43	635	11.3	915	10	US-09-902-853-34	Sequence 34, Appli
44	635	11.3	915	10	US-09-907-824-34	Sequence 34, Appli
45	635	11.3	915	10	US-09-907-841-34	Sequence 34, Appli

ALIGNMENTS

RESULT :
US-09-747-371-2
; Sequence 2, Application US/59747371
; Patent No. US20020006616A1
; GENERAL INFORMATION:
; APPLICANT: Gist, Kurt
; APPLICANT: Mack, David
; TITLE OF INVENTION: NO. US20020006616A1 Methods of Diagnosing Breast Cancer, Comp
; TITLE OF INVENTION: Screening for Breast Cancer Modulators
; FILE REFERENCE: A-69028/DCB/JUD
; CURRENT APPLICATION NUMBER: US/09/747,371
; CURRENT FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: PCT/ US/00/26952
; PRIOR FILING DATE: 2000-03-25
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 2
; LENGTH: 999
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-747-371-2

Query Match		100.0%	Score 5605	DB 9	Length 999
Best Local Similarity		100.0%	Pred No. 0		
Matches 999		Conservative 0	Mismatches 0	Indels 0	Gaps 0
Qy	1 MGVRNRRFGAAWAVLLLLLFPPLLLLAGAVPPGRRAGAGCEVDCAQGLDCHADA 60				
Db	1 MGVRNRRFGAAWAVLLLLLFPPLLLLAGAVPPGRRAGAGCEVDCAQGLDCHADA 60				
Qy	61 LCQNTPTSYKCSCKPQYQGGEGROCEDIDCGNELNGCVHDCNLNIPGNVRCCTCFDGMFLA 120				
Db	61 LCQNTPTSYKCSCKPQYQGGEGROCEDIDCGNELNGCVHDCNLNIPGNVRCCTCFDGMFLA 120				
Qy	121 HDGHNCLDVEDECLNNGGCGQHTCVAVMGSEYECCKEGFELSNQHTCIHRSEGLSCMNK 180				
Db	121 HDGHNCLDVEDECLNNGGCGQHTCVAVMGSEYECCKEGFELSNQHTCIHRSEGLSCMNK 180				

181	Qy	DHGC	SHI	CKE	APRG	S	VACE	CPG	PE	LAK	Q	RDC	IL	T	C	N	H	G	N	G	C	O	H	S	C	D	D	T	A	D	G	P	E	C	S	C	H	241																				
181	Db	DHGC	SHI	CKE	APRG	S	VACE	CPG	PE	LAK	Q	RDC	IL	T	C	N	H	G	N	G	C	O	H	S	C	D	D	T	A	D	G	P	E	C	S	C	H	240																				
241	Qy	POY	K	M	H	T	D	G	R	S	C	L	E	R	E	D	T	V	L	E	V	E	S	N	T	T	S	V	D	G	K	P	U	K	R	R	L	M	E	T	C	A	V	N	G	G	O	R	C	K	D	300						
241	Db	POY	K	M	H	T	D	G	R	S	C	L	E	R	E	D	T	V	L	E	V	E	S	N	T	T	S	V	D	G	K	P	U	K	R	R	L	M	E	T	C	A	V	N	G	G	O	R	C	K	D	300						
301	Qy	STG	V	H	C	S	C	P	V	G	T	L	D	G	K	T	C	K	D	I	E	C	O	T	R	N	G	G	D	H	F	C	K	N	I	V	S	F	D	C	G	K	K	F	L	I	T	D	E	K	360							
301	Db	STG	V	H	C	S	C	P	V	G	T	L	D	G	K	T	C	K	D	I	E	C	O	T	R	N	G	G	D	H	F	C	K	N	I	V	S	F	D	C	G	K	K	F	L	I	T	D	E	K	360							
361	Qy	SQD	V	U	B	E	C	S	L	D	R	T	C	D	H	S	C	I	N	I	P	G	T	F	A	C	A	C	N	R	G	T	L	Y	G	T	H	C	G	T	N	E	C	S	I	N	N	G	C	O	Q	V	C	N	420			
361	Db	SQD	V	U	B	E	C	S	L	D	R	T	C	D	H	S	C	I	N	I	P	G	T	F	A	C	A	C	N	R	G	T	L	Y	G	T	H	C	G	T	N	E	C	S	I	N	N	G	C	O	Q	V	C	N	420			
421	Qy	T	V	G	S	E	C	O	C	H	P	O	Y	K	L	H	N	K	D	C	V	E	V	K	G	L	P	T	S	V	S	P	R	V	S	L	H	C	G	S	G	D	G	C	F	L	R	C	H	S	G	I	E	480				
421	Db	T	V	G	S	E	C	O	C	H	P	O	Y	K	L	H	N	K	D	C	V	E	V	K	G	L	P	T	S	V	S	P	R	V	S	L	H	C	G	S	G	D	G	C	F	L	R	C	H	S	G	I	E	480				
481	Qy	S	D	V	T	T	I	R	T	S	V	T	F	L	N	E	G	K	C	S	L	K	N	A	E	L	P	P	E	G	L	R	P	A	L	P	E	K	H	S	S	V	K	E	S	F	R	Y	N	L	T	C	S	S	G	K	O	540
481	Db	S	D	V	T	T	I	R	T	S	V	T	F	L	N	E	G	K	C	S	L	K	N	A	E	L	P	P	E	G	L	R	P	A	L	P	E	K	H	S	S	V	K	E	S	F	R	Y	N	L	T	C	S	S	G	K	O	540
541	Qy	P	C	A	G	P	R	S	T	P	K	E	M	F	I	T	V	E	F	E	L	N	I	K	E	V	T	A	S	C	D	L	S	C	I	V	K	T	E	K	R	L	K	A	T	R	L	K	A	V	H	E	Q	600				
541	Db	P	C	A	G	P	R	S	T	P	K	E	M	F	I	T	V	E	F	E	L	N	I	K	E	V	T	A	S	C	D	L	S	C	I	V	K	T	E	K	R	L	K	A	T	R	L	K	A	V	H	E	Q	600				
601	Qy	P	H	L	Q	S	G	M	N	L	D	V	A	K	P	P	R	T	S	R	Q	A	E	S	C	G	V	G	G	H	A	E	N	C																								

RESULT 2

```

RESOL 2
US-10-177-293-45
; Sequence 45, Application US/10177293
; Publication No. US2003C124128A1
; GENERAL INFORMATION:
; APPLICANT: Willie, James
; APPLICANT: Giatt, Karen
; APPLICANT: Zhao, Xumei
; APPLICANT: Gannavarpu, Marjula
; APPLICANT: Kanatkar, Subhang
; APPLICANT: Mertens, Maureen
; APPLICANT: Myar, Vic
; APPLICANT: Wang, Youzhen
; APPLICANT: Xu, Yongyao
; APPLICANT: Hoersch, Sebastian
; APPLICANT: Monahan, John

```

```
Db 541 PCAGPSPSTKEMFIVPELETNQKEVTASCDLSIVVRTEKRLKKAIRTLKKAHREQ 600
Qy 601 FHLQLSGMLDVAKKPPRTSEROAESCGVQGHAEHQVSCVCRAGTYDGRARERCILCPNG 660
Db 601 FHLQLSGMLDVAKKPPRTSEROAESCGVQGHAEHQVSCVCRAGTYDGRARERCILCPNG 660
Qy 661 TFQNEEQMTCEPCPRPGNSGALKTPAEANMSECGGLCPQGEVSADGFAPCOLCALGTFQ 720
Db 661 TFQNEEQMTCEPCPRPGNSGALKTPAEANMSECGGLCPQGEVSADGFAPCOLCALGTFQ 720
Qy 721 PEAGRTSCPPCGGLATKQGATSFQDCETRVQCSFGHYNTHHRCIRCPVGTQPEFG 780
Db 721 PEAGRTSCPPCGGLATKQGATSFQDCETRVQCSFGHYNTHHRCIRCPVGTQPEFG 780
Qy 781 KNCVSCPGNTTDFDGSNTITCKNRRCGGLGDFGTGIESPNVPGNYPANTECTWTIN 840
Db 781 KNCVSCPGNTTDFDGSNTITCKNRRCGGLGDFGTGIESPNVPGNYPANTECTWTIN 840
Qy 841 PPKRRILIVVPEIFLPIEDDCGGLYLMRKTSSTSSSVTTTETCQYVERPIAFTSRSKLW 900
Db 841 PPKRRILIVVPEIFLPIEDDCGGLYLMRKTSSTSSSVTTTETCQYVERPIAFTSRSKLW 900
Qy 901 IQKSNESGNSARGFQVPPVYTYDEYQELIEDIVRGRLYASENHQELIKKKLILKALFDV 960
Db 901 IQKSNESGNSARGFQVPPVYTYDEYQELIEDIVRGRLYASENHQELIKKKLILKALFDV 960
Qy 961 LAHPQNYFKYTAQESREMPFRSFRILRLSKVSRFLRYPK 999
Db 961 LAHPQNYFKYTAQESREMPFRSFRILRLSKVSRFLRYPK 999
```

RESULT 3

```
US-10-176-847-50
; Sequence 50, Application US/10176847
; Publication No. US20030068636A1
; GENERAL INFORMATION:
; APPLICANT: Veiby, Pelter Ole
; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR
; IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF BREAST
; TUMORS AND OVARIAN CANCER
; FILE REFERENCE: MRI-039
; CURRENT APPLICATION NUMBER: US/10/176,847
; CURRENT FILING DATE: 2002-06-21
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 50
; LENGTH: 997
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-176-847-50
```

```
Query Match 99.6%; Score 5581; DB 15; Length 997;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 996; Conservative 0; Mismatches 1; Indels 2; Gaps 1;
Qy 1 MGVAGNRPPGAWAV--LLLLPPLLLAGVPPGRGAAGPOEDVDECAQGLDCHADA 60
Db 1 MGVAGNRPPGAWAV--LLLLPPLLLAGVPPGRGAAGPOEDVDECAQGLDCHADA 58
Qy 61 LCQNTPTSYKCSCKPGYGGGROCEDIDCEGNELGVCVHDCLNIPGNVRCCTCFDGMIA 120
Db 59 LCQNTPTSYKCSCKPGYGGGROCEDIDCEGNELGVCVHDCLNIPGNVRCCTCFDGMIA 118
Qy 121 HDGNCILVDDECLNNGGCGHTCNVWMSYECCCKEGFFLSDNQHTCIHRSBEGLSCKMK 180
Db 119 HDGNCILVDDECLNNGGCGHTCNVWMSYECCCKEGFFLSDNQHTCIHRSBEGLSCKMK 178
Qy 181 DHGCSHICKAPRGSVACBPCRFELAKNQDCLTNTGNGGCGQSCDDTDAGPECSCH 240
Db 179 DHGCSHICKAPRGSVACBPCRFELAKNQDCLTNTGNGGCGQSCDDTDAGPECSCH 238
Qy 241 PQYKMTDGRSCLEREDTVLEVTESNTTSVVDGDKVKKRLLMETCAVNGGCDRTCKDT 300
```

```
Db 239 PQYKMTDGRSCLEREDTVLEVTESNTTSVVDGDKVKKRLLMETCAVNGGCDRTCKDT 298
Qy 301 STGVHSCPVGFTLQDLGGCKKQIDECQTRNGGCDHFCKNTVGSFDCGCKKFKLLTDEK 360
Db 299 STGVHSCPVGFTLQDLGGCKKQIDECQTRNGGCDHFCKNTVGSFDCGCKKFKLLTDEK 358
Qy 361 SCQDVDECSLDRTCDHSCINHPGTFACACNRGVTLXGFTHCSDTNECSINNGGCGQCVN 420
Db 359 SCQDVDECSLDRTCDHSCINHPGTFACACNRGVTLXGFTHCSDTNECSINNGGCGQCVN 418
Qy 421 TVGSYECQCHPGKLLHKKKDCVEVKGGLLPTSVPRVSLHCGKSGGDCGFLCHSGIHL 480
Db 419 TVGSYECQCHPGKLLHKKKDCVEVKGGLLPTSVPRVSLHCGKSGGDCGFLCHSGIHL 478
Qy 481 SSVVTTTSTSVTFKJNEGKCSLKNABLFPGLRPPALPEKHSSVKESFRVYVNLTCSSGKV 540
Db 479 SSVVTTTSTSVTFKJNEGKCSLKNABLFPGLRPPALPEKHSSVKESFRVYVNLTCSSGKV 538
Qy 541 PGAPGRPSTKEMFIVPELETNQKEVTASCDLSIVVRTEKRLKKAIRTLKKAHREQ 600
Db 539 PGAPGRPSTKEMFIVPELETNQKEVTASCDLSIVVRTEKRLKKAIRTLKKAHREQ 598
Qy 601 FHLQLSGMLDVAKKPPRTSEROAESCGVQGHAEHQVSCVCRAGTYDGRARERCILCPNG 660
Db 599 FHLQLSGMLDVAKKPPRTSEROAESCGVQGHAEHQVSCVCRAGTYDGRARERCILCPNG 658
Qy 661 TFQNEEQMTCEPCPRPGNSGALKTPAEANMSECGGLCPQGEVSADGFAPCOLCALGTFQ 720
Db 659 TFQNEEQMTCEPCPRPGNSGALKTPAEANMSECGGLCPQGEVSADGFAPCOLCALGTFQ 718
Qy 721 PEAGRTSCPPCGGLATKQGATSFQDCETRVQCSFGHYNTHHRCIRCPVGTQPEFG 780
Db 719 PEAGRTSCPPCGGLATKQGATSFQDCETRVQCSFGHYNTHHRCIRCPVGTQPEFG 778
Qy 781 KNCVSCPGNTTDFDGSNTITCKNRRCGGLGDFGTGIESPNVPGNYPANTECTWTIN 840
Db 779 KNCVSCPGNTTDFDGSNTITCKNRRCGGLGDFGTGIESPNVPGNYPANTECTWTIN 838
Qy 841 PPKRRILIVVPEIFLPIEDDCGGLYLMRKTSSTSSSVTTTETCQYVERPIAFTSRSKLW 900
Db 839 PPKRRILIVVPEIFLPIEDDCGGLYLMRKTSSTSSSVTTTETCQYVERPIAFTSRSKLW 898
Qy 901 IQKSNESGNSARGFQVPPVYTYDEYQELIEDIVRGRLYASENHQELIKKKLILKALFDV 960
Db 899 IQKSNESGNSARGFQVPPVYTYDEYQELIEDIVRGRLYASENHQELIKKKLILKALFDV 958
Qy 961 LAHPQNYFKYTAQESREMPFRSFRILRLSKVSRFLRYPK 999
Db 959 LAHPQNYFKYTAQESREMPFRSFRILRLSKVSRFLRYPK 997
```

RESULT 4

```
US-09-855-824-5
; Sequence 5, Application US/09655824
; Publication No. US20030166348A1
; GENERAL INFORMATION:
; APPLICANT: YAN, Chanhua et al.
; TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,
; NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS, AND
; TITLES OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS,
; FILE REFERENCE: US/09/855,824
; CURRENT APPLICATION NUMBER: US/09/855,824
; CURRENT FILING DATE: 2001-05-16
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 964
; TYPE: PRT
; ORGANISM: Human
US-09-855-824-5
```

Query Match: 96.8%; Score 5428; DB 12; Length 964;

```

Best Local Similarity 100.0%; Pred. No. 0;
Matches 964; Conservative 0; Mismatches 0; Indels 0; Gaps 2;

QY 36 RGRAAGPQEDVDECAQGLDDCHADALCONTPSYKSCPKPYQSGRGCEIDIECGNELN 95
DB 1 RGRAAGPQEDVDECAQGLDDCHADALCONTPSYKSCPKPYQSGRGCEIDIECGNELN 60
QY 96 GCVHDCUNIPGNYSCCTCFDGLMADHGNCLDVDECIENNGGQCHTQCVNNGSYECCK 155
DB 61 GCVHDCUNIPGNYSCCTCFDGLMADHGNCLDVDECIENNGGQCHTQCVNNGSYECCK 120
QY 156 EGGFVSDNQHTCIHRSBGLSCMNKDHGSHICKAPRGSVACRCRPGFELAKQRCIL 215
DB 121 EGGFVSDNQHTCIHRSBGLSCMNKDHGSHICKAPRGSVACRCRPGFELAKQRCIL 180
QY 216 TGNHNGGCGQSCDDTAGPSCSCHPOYKMTDGRSCLEREDTVLEVTESTTSVVDGK 275
DB 181 TGNHNGGCGQSCDDTAGPSCSCHPOYKMTDGRSCLEREDTVLEVTESTTSVVDGK 240
QY 276 RVKRELLMETCAVNVGGDRTCKDTSTGVHSCPVGFTLGLDGKTKCD:DECOFENSGCD 335
DB 241 RVKRELLMETCAVNVGGDRTCKDTSTGVHSCPVGFTLGLDGKTKCD:DECOFENSGCD 300
QY 336 HFCKNIVGSDGCKKGFLLTDEKSCQDVDECSLDRTCDHSCINHPGTAFACACNRGYTL 395
DB 301 HFCKNIVGSDGCKKGFLLTDEKSCQDVDECSLDRTCDHSCINHPGTAFACACNRGYTL 360
QY 336 YGFTHCGDTNECS:INNGCGQCVNTVSYECQHPGYKLHNKDKDVEVKGLLPTSVP 455
DB 361 YGFTHCGDTNECS:INNGCGQCVNTVSYECQHPGYKLHNKDKDVEVKGLLPTSVP 420
QY 456 RVSLHCGKSGGGDGFRLCHSG:HLSSDVTITRTSVTFKLNEGKSLKNAELFPEGRLPA 515
DB 421 RVSLHCGKSGGGDGFRLCHSG:HLSSDVTITRTSVTFKLNEGKSLKNAELFPEGRLPA 480
QY 516 LPEKHSYKESFRYVNLTCSSGKQVPGAPGRSTPKEMFIIVFELETNQKEVTASCDLS 575
DB 481 LPEKHSYKESFRYVNLTCSSGKQVPGAPGRSTPKEMFIIVFELETNQKEVTASCDLS 540
QY 576 CIVKTEKRLKAIITLKAHVREOFHLQSGMNLDAVKKPPRTSEROAESCGVQGHAE 635
DB 541 CIVKTEKRLKAIITLKAHVREOFHLQSGMNLDAVKKPPRTSEROAESCGVQGHAE 600
QY 636 NOCVSCRAGTYDGAERCIICPNGTFONEGOMTCEPCPRPGNSGALKTPAEMNMSCG 695
DB 601 NOCVSCRAGTYDGAERCIICPNGTFONEGOMTCEPCPRPGNSGALKTPAEMNMSCG 660
QY 696 GLCQGEYSADGAFAPQLCALGTQPEAGRTSCFPCCGGGLATKHOGATSFQDCETRVOCS 755
DB 661 GLCQGEYSADGAFAPQLCALGTQPEAGRTSCFPCCGGGLATKHOGATSFQDCETRVOCS 720
QY 756 PGHFYNTTHRCIRCPVGTQPEFGKNNVCSPGNTTDFDGSNTITQCKNRCGGELGD 815
DB 721 PGHFYNTTHRCIRCPVGTQPEFGKNNVCSPGNTTDFDGSNTITQCKNRCGGELGD 780
QY 816 FTGYTESPNYPNPANTECTWTINPPKRRILIVVPEIFLPIEDCCGYLVKRTSSSN 875
DB 781 FTGYTESPNYPNPANTECTWTINPPKRRILIVVPEIFLPIEDCCGYLVKRTSSSN 840
QY 876 SVTTVETCOTYERPIAFTSRSKKLMIOFKNSNGNSARGFQVYVTVYDDEYQELIED:VRD 935
DB 841 SVTTVETCOTYERPIAFTSRSKKLMIOFKNSNGNSARGFQVYVTVYDDEYQELIED:VRD 900
QY 936 GRLYASENHQELIKOKLL:KALFDVLAHPONFYKTAQESREMPFRSIRLLRSKYSRFL 995
DB 901 GRLYASENHQELIKOKLL:KALFDVLAHPONFYKTAQESREMPFRSIRLLRSKYSRFL 960
QY 996 RPYK 999
DB 961 RPYK 964

```

```

US-09-747-371-3
; Sequence 3, Application US/09747371
; Patent No. US2002000661EAI
; GENERAL INFORMATION:
; APPLICANT: Gish, Kurt
; APPLICANT: Mack, David
; TITLE OF INVENTION: No. US2002006616A1el Methods of Diagnosing Breast Cancer. Comp
; TITLE OF INVENTION: Screening for Breast Cancer Modulators
; FILE REFERENCE: A-69028/DJB/USD
; CURRENT APPLICATION NUMBER: US/09/747,371
; CURRENT FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: PCT/ US/02/06952
; PRIOR FILING DATE: 2000-03-15
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patent'n version: 3.0
; SEQ ID NO 3
; LENGTH: 997
; TYPE: PPT
; ORGANISM: Mus sp.
US-09-747-371-3

```

```

Query Match 90.2%; Score 5056; DB 9; Length 997;
Best Local Similarity 89.2%; Pred. No. 0;
Matches 891; Conservative 47; Mismatches 59; Indels 2; Gaps 1;

QY 1 : MGVAGNRPCAAWAV:LLILLAPPLLLAGAVPPGRGAAGFCEDVDECAQGLDDCHADA 60
DB 1 : MGVAGNRPCAAWAV:LLILLAPPLLLAGAVPPGRGAAGFCEDVDECAQGLDDCHADA 59
QY 61 LCQNTPTSYKSCPKPYQSGRGCEIDIECGNELNGGVHDCILNIPNYRCTCFDSEKLA 120
DB 59 LCQNTPTSYKSCPKPYQSGRGCEIDIECGNELNGGVHDCILNIPNYRCTCFDSEKLA 118
QY 121 HDHNCJLDVDECIENNGGQCHTQCVNNGSYECQCKEGFELSDNQHTCIHRSBGLSCNK 180
DB 119 HDHNCJLDVDECIENNGGQCHTQCVNNGSYECQCKEGFELSDNQHTCIHRSBGLSCNK 178
QY 181 DHSCHICKAPRGSVACRCRPGFELAKQRCILTCNHNCGSCQSCDDTAGPSCSCH 240
DB 179 DHSCHICKAPRGSVACRCRPGFELAKQRCILTCNHNCGSCQSCDDTAGPSCSCH 218
QY 241 PQYKMTDGRSCLEREDTVLEVTESTTSVVDGKRVKRLMETCAVNVGGDRTCKDT 300
DB 239 PVRLHADGRSC:LEQGTVLGTESNATSVADGDKRVKRLMETCAVNVGGDRTCKDT 259
QY 301 STGVHSCPVGFTLGLDGKTKCD:DECOFENSGCD:DECOFENSGCD:DECOFENSGCD 360
DB 299 STGVHSCPVGFTLGLDGKTKCD:DECOFENSGCD:DECOFENSGCD:DECOFENSGCD 358
QY 361 SCDDVDECSLDRTCDHSCINHPGTAFACACNRGYTLVGTGTCGDTNECS:INNGCGQCVN 420
DB 359 SCDDVDECSLDRTCDHSCINHPGTAFACACNRGYTLVGTGTCGDTNECS:INNGCGQCVN 418
QY 421 TVGSEYECQHPGYKLHNKDKDVEVKGLLFTSVPRVSLHCGKSGGGDGFRLCHSGIHL 480
DB 419 TVGSEYECQHPGYKLHNKDKDVEVKGLLFTSVPRVSLHCGKSGGGDGFRLCHSGIHL 478
QY 481 SSDVTITRTSVTFKLNEGKSLKNAELFPEGRLPALPEKHSYKESFRYVNLTCSSGKQV 540
DB 479 SSDVTITRTSVTFKLNEGKSLKNAELFPEGRLPALPEKHSYKESFRYVNLTCSSGKQV 538
QY 541 PGAFGSP:PKEMFIIVFELETNQKEVTASCDLS:IVKTEKRLKAIITLKAHVREOF 600
DB 539 PGAFGSP:PKEMFIIVFELETNQKEVTASCDLS:IVKTEKRLKAIITLKAHVREOF 598
QY 601 FHQLSGMNLDAVKKPPRTSEROAESCGVQGHAE:QCVSVCRAGTYDGAERCIICPNNG 660
DB 599 FHQLSGMNLDAVKKPPRTSEROAESCGVQGHAE:QCVSVCRAGTYDGAERCIICPNNG 658
QY 661 TPONEEGQVTCBP:CPHSGS:KATPEANYSBGLQCGPGEYSADGAFAPQLCALGTQ 720
DB 659 TPONEEGQVTCBP:CPHSGS:KATPEANYSBGLQCGPGEYSADGAFAPQLCALGTQ 718

```

```

QY 721 PEAGRTSCPCGGLATKHOGATSFODCETRVOCSPGHFYNTHRCRCPCVGTQPEFG 780
DB 729 PDVGRTSCLSCGGLPTKHGATSFODCETRVOCSPGHFYNTHRCRCPCVGTQPEFG 776
QY 781 KNNVSCPGNTTDFDGSNTITCKNRRCGGELGDTFTGIESPNYPGNYPANTECTWTIN 840
DB 779 KNNVSCPGNTTDFDGSNTITCKNRRCGGELGDTFTGIESPNYPGNYPANSECTWTIN 838
QY 841 PPKRRILLIIVPELFIPIEDDCGDLVMRKTSSTSSNVTYETCOTYERPIATTSKKLM 900
DB 839 PPKRRILLIIVPELFIPIEDDCGDLVMRKTSSTSSNVTYETCOTYERPIATTSKKLM 898
QY 901 IQFKSNEGNSARGFQVPYVYDDEYQELIEDIVDGRLYASENHQELDKKKLIKALFDV 960
DB 899 IQFKSNEGNSARGFQVPYVYDDEYQELIEDIVDGRLYASENHQELDKKKLIKALFDV 958
QY 961 LAHPQNYFKYTAQESREMFPRSRIRLLRSKVSFLRPYK 999
DB 959 LAHPQNYFKYTAQESREMFPRSRIRLLRSKVSFLRPYK 997

RESULT 6
US-09-855-824-6
; Sequence 6, Application: US/09855824
; Publication No. US20030166048A1
; GENERAL INFORMATION:
; APPLICANT: YAN, Chunhua et al.
; TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS, AND
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS, AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: C0001233
; CURRENT APPLICATION NUMBER: US/09/855,824
; CURRENT FILING DATE: 2001-05-16
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 957
; TYPE: PRT
; ORGANISM: Mus muscu.us
US-09-855-824-6

Query Match 88.08; Score 4935; DB 12; Length 957;
Best Local Similarity 90.21; Pred. No. 0;
Matches 862; Conservative 46; Mismatches 48; Indels 0; Gaps 0;

QY 44 EDVDECAQGLDDCHADALCQNTPTSYKSCPKPGYQGEGRQCCEDIDBCGNELNGGVHDCI 103
DB 2 EDVDECAQGLDDCHADALCQNTPTSYKSCPKPGYQGEGRQCCEDIDBCGNELNGGVHDCI 61
QY 104 NIPGNVRCCTCFDGFMLADHGNCLDVBCELNNGGCGHCTVNMVMSYECCKEGRFLSDN 163
DB 62 NIPGNVRCCTCFDGFMLADHGNCLDVBCELNNGGCGHCTVNMVMSYECCKEGRFLSDN 121
QY 164 QHTCIHRSBEGLSOMKDHGCHICKAPRGSVACRCRPGFELAKHQDCILTCNHGNG 223
DB 122 QHTCIHRSBEGLSOMKDHGCHICKAPRGSVACRCRPGFELAKHQDCILTCNHGNG 181
QY 224 QHSCDDTADGPECSCHPQYKMTDGRSCLEREDTVLEVTESNTTSVWGDKRVKRRLLM 283
DB 182 QHSCDDTADGPECSCHPQYKMTDGRSCLEREDTVLEVTESNTTSVWGDKRVKRRLLM 241
QY 284 ETCVAVNGGCDRTCKDTSTGVHSCVPGFTLQDGGTKCIDBCQTRNGGCDHFCNKIYG 343
DB 242 ETCVAVNGGCDRTCKDTSTGVHSCVPGFTLQDGGTKCIDBCQTRNGGCDHFCNKIYG 302
QY 344 SFDGCKKGFLLTDEKSCQDVDECSLDRTCDHSCINHPHTFACACNRGYLYLGFTHCGD 403
DB 302 SFDGCKKGFLLTDEKSCQDVDECSLDRTCDHSCINHPHTFACACNRGYLYLGFTHCGD 362
QY 404 TNECSINNGGCGQVNTVSGYCCQHPGYKHLHNNKQCVKVKGLLPTSVSRVSLHCGK 463
DB 362 TNECSINNGGCGQVNTVSGYCCQHPGYKHLHNNKQCVKVKGLLPTSVSRVSLHCGK 421

```

```

QY 464 SGGDGPCLRCSSGHLSSDVNTIARTSVTFKLNKGKSLKNSLEPEGLREALPEKHSSV 523
DB 422 SGGDGPCLRCSSGHLSSDVNTIARTSVTFKLNKGKSLKNSLEPEGLREALPEKHSSV 481
QY 524 KESFRVNLTCSSGKQVPGAPQRPSTPKEMFTTVTFELETNOKEVYASCDLSCCYKRTK 583
DB 482 KESFRVNLTCSSGKQVPGALGRJNAPKEMFTTVFERETYEKVTASCNLCVYKRTK 541
QY 584 RLKAIKRLAKAVHREDFHLOLSGNLDVAKPPPTSERQAESCGYGHAEHNCQVSCRA 643
DB 542 RLKAIKRLAKAVHREDFHLOLSGNLDVAKPPPTSERQAESCGYGHAEHNCQVSCRA 601
QY 644 GTTYDGARERCILCPNGSTFQNEEGQMTCEPPRPNNGSALKTPEAMNNSCGLCQPEY 703
DB 602 GTTYDGARERCILCPNGSTFQNEEGQMTCEPPRPNNGSALKTPEAMNNSCGLCQPEY 661
QY 704 SAGFAPCQCALGTFTOPBAGRTSCPPCGGGIATKHOGATSFODCETRVOCSPGHFYN 763
DB 662 SAGFAPCQCALGTFTOPBAGRTSCPPCGGGIATKHOGATSFODCETRVOCSPGHFYN 721
QY 764 THRCIRCPVGTQPEFGKNNVSCPGNTTDFDGSNTITCKNRRCGGELGDTFTGIESP 823
DB 722 THRCIRCPVGTQPEFGKNNVSCPGNTTDFDGSNTITCKNRRCGGELGDTFTGIESP 781
QY 824 NYPGNYPANSECTWTINPPPKRRILLIIVPELFIPIEDDCGDLVMRKTSSTSSNVTYETC 883
DB 782 NYPGNYPANSECTWTINPPPKRRILLIIVPELFIPIEDDCGDLVMRKTSSTSSNVTYETC 841
QY 884 QTYERPIATTSKKLMTOFKSNEGNSARGFQVPYVYDDEYQELIEDIVDGRLYASEN 943
DB 842 QTYERPIATTSKKLMTOFKSNEGNSARGFQVPYVYDDEYQELIEDIVDGRLYASEN 901
QY 944 HQBILKDKKLIKALFDVLAPQNYFKYTAQESREMFPRSRIRLLRSKVSFLRPYK 999
DB 902 HQBILKDKKLIKALFDVLAPQNYFKYTAQESREMFPRSRIRLLRSKVSFLRPYK 957

RESULT 7
US-09-855-824-2
; Sequence 2, Application: US/09855824
; Publication No. US20030166048A1
; GENERAL INFORMATION:
; APPLICANT: YAN, Chunhua et al.
; TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS, AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: C0001233
; CURRENT APPLICATION NUMBER: US/09/855,824
; CURRENT FILING DATE: 2001-05-16
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 988
; TYPE: PRT
; ORGANISM: Human
US-09-855-824-2

Query Match 65.6%; Score 3675.5; DB 12; Length 988;
Best Local Similarity 62.4%; Pred. No. 14e-270;
Matches 654; Conservative 113; Mismatches 162; Indels 119; Gaps 14;

QY 10 GAA---WAVLLIILPPILLLAGAVPPGGRGAAG-----POEDVDGCAQGLDDCHADAL 61
DB 2 GAAAVWHLVCLLAL-----GTRGLASGSLPGSDVDVDESGESTDCDCHIDA 49
QY 62 QNTPTSYKSCPKPGYQGEGRQCCEDIDECGNE-LNGGCVHDCIN:PGNYRCCFCGGRFLA 120
DB 50 QNTPTSYKSCPKPGYKGEGRQCCEDIDECNDYNGGCVHCEINIPGNYRCCFCGGRFLA 109
QY 121 HDGHNCLDVBCELNNGGCGHCTVNMVMSYECCKEGRFLSDNHTC:HRSEEGJSCMNK 180
DB 110 HDGHNCLDVBCELNNGGCGHCTVNMVMSYECCKEGRFLSDNHTC:HRSEEGJSCMNK 169

```


181 DHGCSHICAEPRGACBCEPFLAKNORCILTCHNGGCGHSCDDTADGPECSCH 240
170 DHGCAHICRETAKGVACDCRFGFLAQKQKCTLTCHNGGCGHSCDDTADGPECSCH 229
241 POYKXHTDGRSLEREDTVLEVTESTTTSVVDGKRVKRLMETCAVNNGGCORTCKDT 300
230 QKVALHSDGRTCI-----ETCAVNNGGCORTCKDT 259
301 STGVHSCVPVGTFLQDGGTCKDIDECQTRNGGCHFCNIVGSDCCGCKKQKFLITDEK 360
260 ATSVRSCVPVGTFLQDGGTCKDIDECQTRNGGCHFCNIVGSDCCGCKKQKFLITDER 319
361 SCQDVDECSLDRTCDHSCNHPOTFACACNRVYTYGPTHGCGTNECSINNGGCGQCVN 420
320 TCQDIDECSEFERTCHICINSPGSGFCLCHRGYLYGTHGCGDVDECSINSGCQCGVN 379
421 TVGSYECQCHPGYKHLNKKQCVFV-KGLLPTSVSPRVSLHCGKGGGCGFLRC----- 474
380 TKGSYECVCPGPRRLHNGKDCVETGCKLSRAKTSAPRAQLSCSKAGGVESCF-SCPAHTL 439
475 -----HSGIHLS-----SDVTITRTSVTFKLNKCKSL 502
440 FVPDSSENSVYVLCGVPGPOCKALQKNGTSSGLGPGSCSDAPTPIKOKARFKIRDAKCHL 499
503 KNAELFPEGLRPALEKHSVKESPR-----YVNLTCSSGKQVPGAPGRSTPKE 552
500 R-----PHSQARAKETARQPLLDHCHVTFTLKCDSSK--RRGRKSPSKE 544
553 M-FITVEFELETHNQKAVTASCDLSICVIRKTEKRLKAIKATRLKAVHREOFHQLSGCNLD 611
545 VSHITAEFELETHNQKAVTASCDLSICVIRKTEKRLKAIKATRLKAVHREOFHQLSGCNLD 604
612 VAKKPPRTSERQAESCGVQGHAEHQVSCRACTYYDGARERCILCPNTPFONEGQMT 671
605 VAQRPAKALEGG-ACGAGQVLDQSKVACGPGTHFGGELGQCVSCMPGTQDMEGQLSC 663
672 EPCPRGNSGALKTPKAMNMSGGLGCPGEYSADGAPCALGTFQPAAGRTSCFPC 731
664 TPCP---SSDGLGLPGARNVSECGGCGSPGFSADGFKPCQACPVGTQPEGRGTCFPC 720
732 GGLATKHQCATSFQDCETRVQCSGPHFYNTTHRCIRCPVGTQPEFGKNCVSCPGNT 791
721 GGLATKHQCATSFQDCETRVQCSGPHFYNTTHRCIRCPVGTQPEFGKNCVSCPGNT 780
792 TTDFDGSNTITCKNRRCGELGDFGTGYIESPNYGNVYPAFTECWTINPPKRIILVW 851
781 STDFDGSNTITCKNRRCGELGDFGTGYIESPNYGNVYPAFTECWTINPPKRIILVW 840
852 PEIFLPIDDCGDLVWRKTSNNSVTTVETCTYERPIAFTSRSKKLWIOFMSNEGNSA 911
841 PEIFLPIDDCGDLVWRKTSNNSVTTVETCTYERPIAFTSRSKKLWIOFMSNEGNSG 900
912 RGFQVYVYVYDEYQDELIEDIVDRGLYASENHQILKDKKILKALFDVLHPQNYFYKT 971
901 RGFQVYVYVYDEYQDELIEDIVDRGLYASENHQILKDKKILKALFDVLHPQNYFYKT 960
972 AQESREMPFRSIFRLRSKVSRLRPYK 999
961 AQESREMPFRSIFRLRSKVSRLRPYK 988

RESULT 8

US-10-195-142-2
; Sequence 2. Application US/10195:142
; Publication No. US2003036163A1
; GENERAL INFORMATION:
; APPLICANT: Myriad Genetics, Inc.
; APPLICANT: Wettstein, Daniel Albert
; APPLICANT: Mauck, Kimberly A.
; TITLE OF INVENTION: NOVEL PN9826 NUCLEIC ACIDS AND USE THEREOF
; FILE REFERENCE: 1051.02
; CURRENT APPLICATION NUMBER: US/10195:142
; CURRENT FILING DATE: 2002-07-10

PRIOR APPLICATION NUMBER: US 60/304,323
PRIOR FILING DATE: 2001-07-10
NUMBER OF SEQ ID NOS: 39
SOFTWARE: PatentIn version 3.1
SEQ ID NO 2
LENGTH: 988
TYPE: PRT
ORGANISM: Homo sapiens
US-10-195-142-2
Query Match 65.6%; Score 3675.5; DB 15; Length 988;
Best Local Similarity 62.4%; Pred. No. 14e-270;
Matches 654; Conservative 113; Mismatches 162; Indels 115; Gaps 14;
QY 10 GAA---WAVJLLLLLPULLLAGAVPPGGRGAAG-----PQEDVDEACAGLDDCHADAL 61
DB 2 GAAAWHLLCVLLAL-----GTRGLAGGSLPGSVVDVDECEGTDDCHIDAI 49
QY 62 CONTPTSYKSCRPYGGEGROCEDIDECGNE-LJGGCVHDCILNIPNRYRCTCFDGFMLA 120
DB 50 CONTPKSYKCLCPGYKGEKQCEIDECENDYNGGCVHRECINPGNTRCTCFDGFMLA 109
QY 121 HDGHNCLVDDECLENNCGQHTTCVNWVMGYVFCCKEGFFLSPNQHCTCHRSBEG-LSCKMK 180
DB 110 HDGHNCLVDDECLENNCGQHTTCVNWVMGYVFCCKEGFFLSPNQHCTCHRSBEG-LSCKMK 169
QY 181 DHGCSHICAEPRGACBCEPFLAKNORCILTCHNGGCGHSCDDTADGPECSCH 240
DB 170 DHGCAHICRETAKGVACDCRFGFLAQKQKCTLTCHNGGCGHSCDDTADGPECSCH 229
QY 241 POYKXHTDGRSLEREDTVLEVTESTTTSVVDGKRVKRLMETCAVNNGGCORTCKDT 300
DB 230 QKVALHSDGRTCI-----ETCAVNNGGCORTCKDT 259
QY 301 STGVHSCVPVGTFLQDGGTCKDIDECQTRNGGCHFCNIVGSDCCGCKKQKFLITDEK 360
DB 260 ATSVRSCVPVGTFLQDGGTCKDIDECQTRNGGCHFCNIVGSDCCGCKKQKFLITDER 319
QY 361 SCQDVDECSLDRTCDHSCNHPOTFACACNRVYTYGPTHGCGTNECSINNGGCGQCVN 420
DB 320 TCQDIDECSEFERTCHICINSPGSGFCLCHRGYLYGTHGCGDVDECSINSGCQCGVN 379
QY 421 TVGSYECQCHPGYKHLNKKQCVFV-KGLLPTSVSPRVSLHCGKGGGCGFLRC----- 474
DB 380 TKGSYECVCPGPRRLHNGKDCVETGCKLSRAKTSAPRAQLSCSKAGGVESCF-SCPAHTL 439
QY 475 -----HSGIHLS-----SDVTITRTSVTFKLNKCKSL 502
DB 440 FVPDSSENSVYVLCGVPGPOCKALQKNGTSSGLGPGSCSDAPTPIKOKARFKIRDAKCHL 499
QY 503 KNAELFPEGLRPALEKHSVKESPR-----YVNLTCSSGKQVPGAPGRSTPKE 552
DB 500 R-----PHSQARAKETARQPLLDHCHVTFTLKCDSSK--RRGRKSPSKE 544
QY 553 M-FITVEFELETHNQKAVTASCDLSICVIRKTEKRLKAIKATRLKAVHREOFHQLSGCNLD 611
DB 545 VSHITAEFELETHNQKAVTASCDLSICVIRKTEKRLKAIKATRLKAVHREOFHQLSGCNLD 604
QY 612 VAKKPPRTSERQAESCGVQGHAEHQVSCRACTYYDGARERCILCPNTPFONEGQMT 671
DB 605 VAQRPAKALEGG-ACGAGQVLDQSKVACGPGTHFGGELGQCVSCMPGTQDMEGQLSC 663
QY 672 EPCPRGNSGALKTPKAMNMSGGLGCPGEYSADGAPCALGTFQPAAGRTSCFPC 731
DB 664 TPCP---SSDGLGLPGARNVSECGGCGSPGFSADGFKPCQACPVGTQPEGRGTCFPC 720
QY 732 GGLATKHQCATSFQDCETRVQCSGPHFYNTTHRCIRCPVGTQPEFGKNCVSCPGNT 791
DB 721 GGLATKHQCATSFQDCETRVQCSGPHFYNTTHRCIRCPVGTQPEFGKNCVSCPGNT 780
QY 792 TTDFDGSNTITCKNRRCGELGDFGTGYIESPNYGNVYPAFTECWTINPPKRIILVW 851
DB 781 STDFDGSNTITCKNRRCGELGDFGTGYIESPNYGNVYPAFTECWTINPPKRIILVW 840
QY 852 PEIFLPIDDCGDLVWRKTSNNSVTTVETCTYERPIAFTSRSKKLWIOFMSNEGNSA 911
DB 841 PEIFLPIDDCGDLVWRKTSNNSVTTVETCTYERPIAFTSRSKKLWIOFMSNEGNSG 900
QY 912 RGFQVYVYVYDEYQDELIEDIVDRGLYASENHQILKDKKILKALFDVLHPQNYFYKT 971
DB 901 RGFQVYVYVYDEYQDELIEDIVDRGLYASENHQILKDKKILKALFDVLHPQNYFYKT 960
QY 972 AQESREMPFRSIFRLRSKVSRLRPYK 999
DB 961 AQESREMPFRSIFRLRSKVSRLRPYK 988

QY 852 PEIPLIEDDCGVLMNRKTSNSSVTTTCTQYERPIAFTSRKLMWQKSNRNSA 912
DB 841 PEIPLIEDDCGVLMNRKTSASPTSTTTCTQYERPIAFTSRKLMWQKSNRNSG 900
QY 912 RGFOVPVTVYDDEYQELIEDIVDRGRLYASENHQELKDKKLIKALFDVLAHFQNYFKYT 971
DB 901 KGFQVPVTVYDDEYQELIEDIVDRGRLYASENHQELKDKKLIKALFDVLAHFQNYFKYT 960
QY 972 AQESREMPFRSFIKLLRSKVSRLRPYK 999
DB 961 AQESKEMFPRSFILKLLRSKVSRLRPYK 988

RESULT 9

US-09-898-570-14
; Sequence 14, Application US/09898570
; Patent No. US20020123612A1
; GENERAL INFORMATION:
; APPLICANT: GERLACH, VALERIE L.
; APPLICANT: ELLERMAN, KAREN
; APPLICANT: MACDOUGALL, JOHN R.
; APPLICANT: SMITHSON, GLENDA
; TITLE OF INVENTION: NOVEL HUMAN PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND
; TITLE OF INVENTION: METHODS OF USING THE SAME
; FILE REFERENCE: 15966-776CIP
; CURRENT APPLICATION NUMBER: US/09/898,570
; CURRENT FILING DATE: 2001-07-03
; PRIOR APPLICATION NUMBER: 60/198,293
; PRIOR FILING DATE: 2000-04-19
; PRIOR APPLICATION NUMBER: 60/198,645
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: 60/210,809
; PRIOR FILING DATE: 2000-06-09
; PRIOR APPLICATION NUMBER: 60/199,476
; PRIOR FILING DATE: 2000-04-26
; PRIOR APPLICATION NUMBER: 60/200,035
; PRIOR FILING DATE: 2000-04-26
; PRIOR APPLICATION NUMBER: 60/224,610
; PRIOR FILING DATE: 2000-08-11
; PRIOR APPLICATION NUMBER: 60/200,024
; PRIOR FILING DATE: 2000-04-26
; PRIOR APPLICATION NUMBER: 60/199,880
; PRIOR FILING DATE: 2000-04-26
; PRIOR APPLICATION NUMBER: 60/218,591
; PRIOR FILING DATE: 2000-07-17
; PRIOR APPLICATION NUMBER: 60/271,814
; PRIOR FILING DATE: 2001-02-27
; PRIOR APPLICATION NUMBER: 60/215,855
; PRIOR FILING DATE: 2000-07-03
; PRIOR APPLICATION NUMBER: 09/839,446
; PRIOR FILING DATE: 2001-04-19
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 14
; LENGTH: 974
; TYPE: PRT
; ORGANISM: Unknown Organism

FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: POLYX
; OTHER INFORMATION: 297832_B.1
US-09-898-570-14

Query Match 61.0%; Score 3417.5; DB 10; Length 974;
Best Local Similarity 60.1%; Pred. No. 5.7e-251;
Matches 596; Conservative 140; Mismatches 219; Indels 37; Gaps 9;

QY 22 LPPLLILAGVPPGGRAGQEDVDECAQGLDCHADALCONTPTSYKSCXKPGYQEG 81
DB 6 VEGCLLVLLVHARAQYKAAQDVDECVGTGDNCHDAICONTPSYKICKSGYTG 65
QY 92 RCECIDECEGNEGCGVHDCINIPGNRYCTCFDGFMLAHGHNCLDVBCELENNGSCQH 141

DE 66 KHCKVDDECEBERNAGCVHDCVNI:PGNYRCTCYDGFHLADHGNCLDVBCELENNGSCQH 125
QY 142 TCNVNMGSYCCCKBESFFILSDNCHTCIHRSEBGLSCMNKDHGCSHICKAPRGSVACER 201
DB 126 SCNVNMGSYCHCREGFFISDNCHTCIHRSEBGLSCMNKDHGCSHICKAPRGSVACER 185
QY 202 PGPELANQSCCLTGNHNGGCGHSCDDTADQPECSCHPOYKMTDGRSCLERDQTVLE 261
DB 186 PGPELANQSCCLTGNHNGGCGHSCDDTADQPECSCHPOYKMTDGRSCLERDQTVLE 237
QY 262 VTESNTTSVVDGSKVRKRLJ-----LVETCAVNNGGCCTCTCTSTGVHCSQVGFLOL 316
DB 238 -----GERRLECHPTCAVSNETCAVNNGGCDSKCHDAATGVHCTCPVGFMLQP 286
QY 317 DGTCKDIDEOTRNGGCHFCNIVGSEPCCKKGFLLTDEKSCODVDESLDRTCOH 376
DB 287 DRKCKDI:DECLNNGGCHICRNTVGSFECCKGKYLINERNCCDIDESCDFRICH 346
QY 377 SCINHPGTFAACNRYGTYLGFTHCGDTNECSINNGGCGQCVNTVGSYECOCCHPOY-KJ 435
DB 347 ICVNTPGSFQCLCHRGYLLYGITHCGDVDECSINRGCGRFGCINTPGSYOCTCPAQOGR 406
QY 436 HWKKKOCVEVKGILLPTSVSPVSLHCKSGGCGCFRCHSGHL--SSDVTTIRTSVTF 493
DB 407 HWNGKOCTEPLKCGSPGASKAMLSNRSCKKDTCAITCPSRARFLPEAAVLISKORASF 466
QY 494 KLNKGC--SLKNAELFPEGLRPALEPKHSSVKESPRYVNLTCSSGKQVPGAPRSTPK 551
DB 467 KIMDAKRLH:RNKGTTEAGRTTGGGAPCSCQVTFILKCDSSRKSKGRAPRTPPCK 526
QY 552 EMP+TVEFELETKQEVNTASCDLSIVKRTKRLKAIKRTLRKAVHRSQFHLQSLMNL 610
DB 527 EVTRLTLEAEVPAEETTASQGLPCLRQMERRLKGSLKLRKSLNQRELLRLAGLDY 586
QY 611 DVAKKPRTSERQA---ESCGVGGAHQVSCVCRAGTYVDGARERCILCPNGTFCWESG 667
DB 587 EAHKPGLVAGERAEPVESCRRPQHRAGTKVSCPCQGTTHGQTBQCVPCFAGTFQEREG 646
QY 668 QMTCEPCPRFGNSGALKITPEANMSECGGLCQPGEYSADGFAPCQLCALGTTPQEAGR 727
DB 647 QLSCLCPGSDAHGFL--GATNVITCAGOCPPGQSVDFKPCQCPRGTYQPEAGR 703
QY 728 CFPGGGLATKHQNTSFQDCETRVQSGPHFNTTHRCIRCPVGTQYQPEGKNVCSC 787
DB 704 CFPGGGLATKHQNTSFQDCETRVQSGPHFNTTHRCIRCPVGTQYQPEGKNVCSC 763
QY 788 PGNITTFDGSNTITCKNRRCGELGDTGYIESPNYPONYPANTECTWTNPPPKRI 847
DB 764 PGNITTFDGSNTITCKNRRCGELGDTGYIESPNYPONYPANTECTWTNPPPKRI 823
QY 848 LIVVPEIFLPIEDDCGVLMNRKTSNSSVTTTCTQYERPIAFTSRKLMWQKSNRNSA 907
DB 824 LIVVPEIFLPIEDDCGVLMNRKTSNSSVTTTCTQYERPIAFTSRKLMWQKSNRNSA 893
QY 908 GNSARGQVYVTVYDDEYQELIEDIVDRGRLYASENHQELKDKKLIKALFDVLAHFQNY 967
DB 884 ANSARGQVYVTVYDDEYQELIEDIVDRGRLYASENHQELKDKKLIKALFDVLAHFQNY 943
QY 968 FKYTAQESREMPFRSFIKLLRSKVSRLRPYK 999
DB 944 FKYT-EKHKEMLPKSFILKLLRSKVSRLRPYK 974
RESULT 10
US-C9-839-446-14
; Sequence 14, Application US/09839446
; Publication No. US20030050232A1
; GENERAL INFORMATION:
; APPLICANT: GERLACH, VALERIE L.
; APPLICANT: ELLERMAN, KAREN
; APPLICANT: MACDOUGALL, JOHN R.
; APPLICANT: SMITHSON, GLENDA
; TITLE OF INVENTION: NOVEL HUMAN PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND

TITLE OF INVENTION: METHODS OF USING THE SAME

FILE REFERENCE: 15966-776

CURRENT APPLICATION NUMBER: US/09/839,446

CURRENT FILING DATE: 2001-04-19

PRIOR APPLICATION NUMBER: 60/198,293

PRIOR FILING DATE: 2000-04-19

PRIOR APPLICATION NUMBER: 60/198,645

PRIOR FILING DATE: 2000-04-20

PRIOR APPLICATION NUMBER: 60/210,909

PRIOR FILING DATE: 2000-06-09

PRIOR APPLICATION NUMBER: 60/199,476

PRIOR FILING DATE: 2000-04-26

PRIOR APPLICATION NUMBER: 60/200,025

PRIOR FILING DATE: 2000-04-26

PRIOR APPLICATION NUMBER: 60/224,610

PRIOR FILING DATE: 2000-08-11

PRIOR APPLICATION NUMBER: 60/200,024

PRIOR FILING DATE: 2000-04-26

PRIOR APPLICATION NUMBER: 60/199,880

PRIOR FILING DATE: 2000-04-26

PRIOR APPLICATION NUMBER: 60/218,591

PRIOR FILING DATE: 2000-07-17

PRIOR APPLICATION NUMBER: 60/271,814

PRIOR FILING DATE: 2001-02-27

NUMBER OF SEQ ID NOS: 49

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 14

LENGTH: 974

TYPE: PRT

ORGANISM: Unknown Organism

FEATURE:

OTHER INFORMATION: Description of Unknown Organism: POLYX

OTHER INFORMATION: 297832_B_1

US-09-839-446-14

Query Match 61.0%; Score 3477.5; DB 11; Length 974;

Best Local Similarity 60.4%; Pred. No. 5,7e-251;

Matches 596; Conservative 140; Mismatches 219; Indels 37; Gaps 9;

22 LPELLLAGAVPPGRRAAGPQSDVDECAQGLDDCAHALCQNTPTSYKSCSPGYGSG 91
 6 VPGCLLVLLVHARAAGYSKAAQDVDECEGTGDKHDALCQNTPTSYKSCSPGYGSG 65
 92 RQCEIDIDECNEINSGCCVDCINTPGNTPTCTFDGFMALHNGNCLDVBCELENNSSCGH 141
 66 KHCVDVDECEENACVHDCVMTDQVPTCTCYDGFHLAGHNCCLDVBCELENNSSCGH 125
 142 TCYVVGSGYCCCKEAGFFLSDNQHCTHSEEGSLCKMKKHGSHCEKAPRGSVACGR 201
 126 SCYVVGSGYCHCRBFFLSDNQHCTHSEEGSLCKMKKHGSHCEKAPRGSVACGR 185
 202 PGFELANQDCLITLTHNGNCGSCDVTADGPECSCHPYKXHTDGRCLEREDTVLE 261
 186 PGFELANQDCLITLTHNGNCGSCDVTADGPECSCHPYKXHTDGRCLEREDTVLE 237
 262 VTESNTSVVDGDKVKKRL-----LME7CAVNNGGCCTKCKDSTGVMSCTVGTLLQ 316
 238 -----GSRLECHIPTQAVSNE7CAVNNGGCCKDHAATGHTCGVGNLQ 286
 317 DGTCKXIDECQTRNNGSCDHPKNTVGSFDCGCKGFFLLIDESSCCVDECSLPTTCH 376
 287 DRTCKXIDECQTRNNGSCDHPKNTVGSFDCGCKGFFLLIDESSCCVDECSLPTTCH 346
 377 SCINHPGTFACACNRYLYLFTGCGDTNCSINNNGSCQVQVAVTVSGYEQCHPGY-RL 435
 347 ICVTPDSFGLCHRYLYLYGITHCGDVCEKINGGRCPCINTPGSYCTCFAGCGRL 406
 436 HMKKDCVEYKGLPTVSFVSLHCGSGGGGCFHCHGSHL-----SSVTIIRSVYE 493
 407 HMKKDCVEYKGLPTVSFVSLHCGSGGGGCFHCHGSHL-----SSVTIIRSVYE 466
 494 KLEGGK--SLKNAELFPGCLRPALPEKHSYKESFRVWLTSSGSGQVGFAGRPSTPK 551

DB 467 KIKCAKCRLLHLRNKCKCEAGRTTGGQAPCECVTFILKCDSSRRKCKGRAPTGGK 526
 CY 552 EME-ITVEPELETNKGKVTASCDLSCVKTERTRLPKAIFRLKXAVHCEFLDLSGNN 610
 DB 527 EVTRLITLEAVEVAEETASCGPLCIROMERRLKGSJKYLRKINDORFLRLAGLDY 586
 CY 611 DVAKKPRPTSEQA---ESGVOGHAENQVSCRASTYVSGAREPCLICNPTFQNEEG 667
 DB 587 EYAKRPGVAGERRAPNESCPGGRAGTKVSCPGCTVYHJQTEGCVCPAGTQREG 646
 CY 668 CMTCEPCEPFGSGALTKTPEAMNSECGLCPGEXSARGFAPCVCAALGTQEPAGRTS 727
 DB 647 QUSCOLCPGSCAHGL---CATVTTAGCCPFGHSDVGFRCQCPGPTGTPAGRTL 703
 CY 728 CFFGSGSLATXHQATSFQDCETPVGSCPGHFNVTTHRCIRCPVATYAPFGKKNVSC 787
 DB 734 CFFGSGSLATXHQATSFQDCETPVGSCPGHFNVTTHRCIRCPVATYAPFGKKNVSC 763
 CY 765 FGNTTDPPDGSNTIQCCKNRRCGGLDFTGYTESPVGNYPANTECTWTINPPKRI 847
 DB 764 FGNTTDPPDGSNTIQCCKNRRCGGLDFTGYTESPVGNYPANTECTWTINPPKRI 823
 CY 848 LIVPEITLPIEDCCDYLWKRKTSNSVTTETCTERPAPTSKXKXLCFKNE 907
 DB 824 LIVPEITLPIEDCCDYLWKRKTSNSVTTETCTERPAPTSKXKXLCFKNE 883
 CY 908 CNAAGFQVAVTYDEEYQSLIEDIVDGRVYASEHCEILDKKILKALFVLAHPONY 967
 DB 884 ANSARGFQVAVTYDEEYQSLIEDIVDGRVYASEHCEILDKKILKALFVLAHPONY 943
 CY 968 EKYTAQSRSEMPRFRILRLSKYSRFLRZYK 999
 DB 944 EKYTAQSRSEMPRFRILRLSKYSRFLRZYK 974

RESULT 11
 US-10-239-663-36
 Sequence 36, Application US/10239663
 Publication No. US20030139572A1
 GENERAL INFORMATION:
 APPLICANT: Agniwal, Pankaj
 APPLICANT: Murdoch, Paul R.
 APPLICANT: Rizvi, Saliya, K.
 APPLICANT: Smith, Randall, F.
 APPLICANT: Xiang, Zhaoying
 APPLICANT: Kadrick, Karen
 TITLE OF INVENTION: NOVEL COMPOUNDS
 FILE REFERENCE: GP50018
 CURRENT APPLICATION NUMBER: US/10/239,663
 PRIOR FILING DATE: 2002-03-24
 PRIOR APPLICATION NUMBER: PCT/US01/09226
 PRIOR FILING DATE: 2001-03-22
 PRIOR APPLICATION NUMBER: 60/192,158
 PRIOR FILING DATE: 2000-03-24
 PRIOR APPLICATION NUMBER: 60/192,668
 PRIOR FILING DATE: 2000-03-27
 PRIOR APPLICATION NUMBER: 60/203,166
 PRIOR FILING DATE: 2000-04-27
 NUMBER OF SEQ ID NOS: 65
 SOFTWARE: PaeSite for Windows Version 3.0
 SEQ ID NO 36
 LENGTH: 993
 TYPE: PRT
 ORGANISM: Homo sapiens
 US-10-239-663-36

Query Match: 60.4%; Score 3397; DB 12; Length 993;
 Best Local Similarity: 56.3%; Pred. No. 2.1e-249;
 Matches 556; Conservative 136; Mismatches 212; Indels 78; Gaps 9;

22 LPELLLAGAVPPGRRAAGPQSDVDECAQGLDDCAHALCQNTPTSYKSCSPGYGSG 81
 6 VPGCLLVLLVHARAAGYSKAAQDVDECEGTGDKHDALCQNTPTSYKSCSPGYGSG 65


```

QY 377 SCINHPGTFACACNAGTYLFTGTHGCDNEGS:ANGCCQYCVANTVSGYECQCHPGY-KL 435
DB 347 ICVNTPGSFQCLCHRGYLLYSITGCGDVECS:INNGGRCFGCINTPGSYCCTCPAGGGRLL 406
QY 436 HMKKDCVEVKGLLFTSVSPRYS:LCGSGGGGDCGCFIAC----- 474
DB 407 HMKKDCTEPLKCGSSPGASAKAMLS:CNBSGKKDTALTCPSPARFLPSENGTVCSTP 446
QY 475 -----HSGIHLSSD-----VTTIRTSVTFKLNEGKC--SLKNAEAFPEGJRFAL 516
DB 467 SPRAAPARAGHNGNSTNSHCHEAVALSIKQASASKIDANCRLHLRNKGTKEAGRTG 526
QY 517 PEKSSSVKESFRYYNLTGSSGKQVPGAPGRSPSTPKMF--ITVEFELETOKEVTAACDLS 575
DB 527 PGAPAPCECCQVTFIHLKCDSSRKGRARATPPGKEVTRLTLELAEVRAEETTAACGLP 586
QY 576 CIVKRTKRLKRAIRTC:LRKAVHREDFHLQLSGMNLJYAKKPPR:SERQA---ESGCVGCG 632
DB 587 CLRQMERRLKGSJMLRKSIINQCFLLRLAGLDYE:LAHKGLVAGGERABMEBCRPGCH 646
QY 633 HAENOCVSCRACTVYDGAARERCILCPNGTFQNEEGQMTCEPCPRPGNSGALKTPEAMKXS 692
DB 647 RAGTKCVSCPGCTVYHGGTEQCVCPCAGTFOREBQCLSDCLCPSSDAHGPL---GATVVT 703
QY 633 ECGGLCQPGEVASADGFAPRCULGALGTFQPEBAGRISCFPGCGG:ATKQGSATSFQDCETRY 752
DB 764 TCAGCCPPGQHSVDGFKPCQCP:PRGTYQPEAGRTLCFPGCGGLTCKHGCAISFQDCDCTKY 763
QY 753 QCSPGHFYNTTTRGTCRPGVGTQPEFGKKNQVSG:RGNTTDFQGSNTITQCKKRRGCGE 812
DB 764 QCSPGHYNTSTIRCTRCAMGSIYQDFPRQNFCSRCPGNTSTCPDGSITSVAQCKNRQCGE 823
QY 813 LGDTGYIESPNYPNPANTECTWTINPPKRR:LLIVPELFLPIEDCGDYLMKXTS 872
DB 824 LGFTGYIESPNYPNPAGVCECININPPKRL:LIIVPELFLPSEDECCVLYMRKNS 883
QY 873 SSNSVTTTETCOTYRPAFTSRSKKMLIQFKSNKNGSARGQVYVYDDEYQELLBDI 932
DB 884 SPSSITTYETCOTYRPAFTARSKMLINFSTSPANSARGQIYVYVYDDEYQELLBDI 943
QY 933 VDDGRLVASENHQELLKXKKIKALFDVLAHPQNYFKYTAQESREMPFRSFLRLRSKVS 992
DB 944 VDDGRLVASENHQELLKXKKIKALFEVLAHQNYFKIT--EKMEXMLPKSFKILRSKVS 1002
QY 993 RLRLPYK 999
DB 1003 SFLRLPYK 1009

```

```

PRIOR FILING DATE: 2000-04-26
PRIOR APPLICATION NUMBER: 60/224,610
PRIOR FILING DATE: 2000-08-11
PRIOR APPLICATION NUMBER: 60/200,024
PRIOR FILING DATE: 2000-04-26
PRIOR APPLICATION NUMBER: 60/199,880
PRIOR FILING DATE: 2000-04-26
PRIOR APPLICATION NUMBER: 60/218,591
PRIOR FILING DATE: 2000-07-17
PRIOR APPLICATION NUMBER: 60/271,814
PRIOR FILING DATE: 2001-02-27
NUMBER OF SEQ ID NOS: 49
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO: 16
LENGTH: 1009
TYPE: PRT
ORGANISM: Unknown Organism
FEATURE:
OTHER INFORMATION: Description of Unknown Organism: POLYX
OTHER INFORMATION: G355056-04
US-09-839-446-16

Query Match      60.6%  Score 33941  DE 11  Length 1009:
Best Local Similarity 58.1%  Pred. No. 3 6e-249:
Matches 597; Conservative 142; Mismatches 216; Indels 72; Gaps 10:

QY 22 LPELLLLAGAVPFGRRAPAPCEDEVEDCAQGLDDCHADALCONTFSTYKSGCKFYQGS 81
DB 6 VPEGLCLVLLVHARAQYKAAQDVDECEGTGDNCHIDAICQNTPRSYKICIKSGYTGDS 65
QY 82 ROCEJIDECENE:NGCCVHDCUNIPENYECTGDSKX:LAHDGNCICDVDECCJENNGGCH 141
DB 66 RKCKDVDECEBENACVADCVNIPENYCTCYDGFHLHADGNCICDVDECCJENNGGCH 125
QY 142 TCYNMAGSYECCKEGFPFLSDQNTTIFHSEEG:LSGMNKHGCSHICKAPRGSVACER 201
DB 126 SCVNMAGSYCHREBGFLLSDNQTIOREBEMNMNKHGCAHICRETPRGK:ACER 165
QY 202 PGFELAKXPRDCLITNHNAGGCGHSCDQTDABPECSCHPQYKMTDGSGLEREDTVJE 261
DB 186 PGFELTNQDRCKLTCYNAGGCGHSCDQTECGPRGCHIKFVLTIDGKTCI----- 217
QY 262 VTSSNTSVVDGKXRYKRR:-----LMEICAVNNGGCDPTCKDTSTGYHCSGCVGFTLOL 316
DB 238 -----GRRLEQHTPTQAVSNBTOAVNNGGCDSCKGRALATGHTCGVGYLOP 286
QY 317 DGKTCXIDECQTRNGGCDHFCXK:VGSFDCGCKGFKLLTDEKSCQDYDECSJDRTCDH 376
DB 287 DRKTCXIDECR:AMNGGCHICNTVGSFECSCKGYKLLINRNCQDIDERFDRTCDH 346
QY 377 SCINHPGTFACACNAGTYLFTGTHGCDNEGS:ANGCCQYCVANTVSGYECQCHPGY-KL 435
DB 347 ICVNTPGSFQCLCHRGYLLYSITGCGDVECS:INNGGRCFGCINTPGSYCCTCPAGGGRLL 406
QY 436 HMKKDCVEVKGLLFTSVSPRYS:LCGSGSGSGGCGCFIAC----- 474
DB 407 HMKKDCTEPLKCGSSPGASAKAMLS:CNBSGKKDTALTCPSPARFLPSENGTVCSTP 446
QY 475 -----HSGIHLSSD-----VTTIRTSVTFKLNEGKC--S:KNAELFPEGJRFAL 516
DB 467 SPRAAPARAGHNGNSTNSHCHEAVALSIKQASASKIDANCRLHLRNKGTKEAGRTG 526
QY 517 PEKSSSVKESFRYYNLTGSSGKQVPGAPGRSPSTPKMF--ITVEFELETOKEVTAACDLS 575
DB 527 PGAPAPCECCQVTFIHLKCDSSRKGRARATPPGKEVTRLTLELAEVRAEETTAACGLP 586
QY 576 CIVKRTKRLKRAIRTC:LRKAVHREDFHLQLSGMNLJYAKKPPR:SERQA---ESGCVGCG 632
DB 587 CLRQMERRLKGSJMLRKSIINQCFLLRLAGLDYE:LAHKGLVAGGERABMEBCRPGCH 646
QY 633 HAENOCVSCRACTVYDGAARERCILCPNGTFQNEEGQVTCPCPRPGNSGALKTPEAMKXS 692
DB 647 RAGTKCVSCPGCTVYHGGTEQCVCPCAGTFOREBQCLSDCLCPSSDAHGPL---GATVVT 703

```

QY 693 ECGGLCCPGEVSADGAPACQCALGTQPEAGRTSCPPCGGLATKHQAGTSPODESTRV 752
 DB 704 TCAQCPCPGGSHSVGPFRCQPCPRGTQPEAGRTLCPPCGGLTKKAEGLSPQDDCTKY 763
 QY 753 QCSGSHFYNTTHHCIRCPVGTQPEFGKNNVCSPNNTTTFDGSNTITQCKXPRGCE 812
 DB 764 QCSGSHYNTSIHHCIRKAMGSYQPDPRQNCRCRPNSTDPDGSISVAGCKNRGCGE 823
 QY 813 LGDTGYIESPNYNGNYPANTECTWTNPPKRRILIVPEILPIEDDCGDLVWKRKS 872
 DB 824 LGSEFGYIESPNYNGNYPAGVEICWINPPKRRILIVPEILPSEDECGDLVWKRKS 883
 QY 873 SSNSVTYETQTERPIAFTSRKSLMIQKSNEGSARGQVPPYVTDYQELLEDI 932
 DB 884 SPSSITVETQTERPIAFTSRKSLMIQKSNEGSARGQVPPYVTDYQELLEDI 943
 QY 933 VRDRLYASENHQELIKDKLIKALFDVLAPQVFFYTAQESREMPFRAFIILSKVS 992
 DB 944 VRDRLYASENHQELIKDKLIKALFEVLAPQVFFYTAQESREMPFRAFIILSKVS 1002
 QY 993 RFLRPYK 995
 DB 1003 SFLRPYK 1009

RESULT 14

US-10-239-663-58
 Sequence 58, Application US/10239663
 Publication No. US20030139572A1
 GENERAL INFORMATION:
 APPLICANT: Agarwal, Pankaj
 APPLICANT: Murdock, Paul R.
 APPLICANT: Rizvi, Safia, K.
 APPLICANT: Smith, Randall, F.
 APPLICANT: Xiang, Zhaoying
 APPLICANT: Kabnick, Karen
 TITLE OF INVENTION: NOVEL COMPOUNDS
 FILE REFERENCE: GP50018
 CURRENT APPLICATION NUMBER: US/10/239, 663
 CURRENT FILING DATE: 2002-09-24
 PRIOR APPLICATION NUMBER: PCT/US01/09226
 PRIOR FILING DATE: 2001-03-22
 PRIOR APPLICATION NUMBER: 66/192,158
 PRIOR FILING DATE: 2000-03-26
 PRIOR APPLICATION NUMBER: 66/192,668
 PRIOR FILING DATE: 2000-03-27
 PRIOR APPLICATION NUMBER: 60/200,166
 PRIOR FILING DATE: 2000-04-27
 NUMBER OF SEQ ID NOS: 66
 SOFTWARE: FASTSEQ for Windows Version 3.0
 SEQ ID NO 58
 LENGTH: 964
 TYPE: PRT
 ORGANISM: Homo sapiens
 US-10-239-663-58

Query Match

58.7%: Score 3290.5; DB 12: Length 964;
 Best Local Similarity 59.7%: Fred. No. 2.5e-241;
 Matches 580: Conservative 110; Mismatches 162; Indels 119; Gaps 14;

QY 10 GAA---AAVLLLLLLPPLLALAGAVPPGRGAAG-----PQEVDECAQGLDQCHADAL 61
 DB 2 GAAVRAHMLCVLLAL-----GTRGLAGGSGSLPSSVVDDESSESTDCHIDAI 49
 QY 62 CQNTPTSYKSCCKPGYQGEGRQCEDIDECGNE-LNGCVHDLCLNPGNYRCTCFDGMFLA 120
 DB 50 CQNTPTSYKCLCKPGYKGEGRQCEDIDECNDYVNGCVHCECINPGNYRCTCFDGMFLA 109
 QY 121 HDGHNCLDVDECLENNCGQCHTGVVWMSYEGCCCEGFFLSNDHCTCHRESEGLSCMNK 180
 DB 110 HDGHNCLDVDECDNNGCGQCHTGVVWMSYEGCCCEGFFLSNDHCTCHRESEGLSCMNK 169

QY 181 DECSHICKEAPRGSVACERPGFELANQPDCLTCNHNNGGCGHSCDDTADPECSCH 240
 DB 170 DHCCHICRETPKGVACDCRPGFLANQPDCLTCNHNNGGCGHSCDDTADPECSCH 229
 QY 241 PQYKMTDGRSCLEEDTGLEVTESNTSVVDGRKVRKRLMETCAVNNGCCDPTCKD 300
 DB 230 QYALHSDGRTCI-----ETCAVNNGCCDPTCKD 259
 QY 301 STGVHSCGPVGTJLDBKTCCKJDEQOTRNGSCDHFGKNVGSFDDCCCKGFKLLTDEK 360
 DB 260 ATGVHSCGPVGTJLDBKTCCKJDEQOTRNGSCDHFGKNVGSFDDCCCKGFKLLTDEK 319
 QY 361 SCQVDDECSLDRTCHSCINHPGTACACNNGYTLGFTTHGDTNECS-INNGCCQVGVN 420
 DB 320 TQODIDEGSFERTCHICINSNGSTQCCHNGYTLGFTTHGDTNECS-INNGCCQVGVN 379
 QY 421 TVGSYECQHPGYLHMKKQCEV-KQLPTSVSPVSLHCGSGGSDGCFRLC----- 474
 DB 380 TVGSYECQHPGYLHMKKQCEV-KQLPTSVSPVSLHCGSGGSDGCFRLC----- 439
 QY 475 -----HSGHLS-----SDVTIRTSVTFKLNKCGSL 502
 DB 440 FVPSNENSVYVSCVPGPQKALQKRNQTSGLGFGSCDAPTPKQYARFIRIRAKHL 499
 QY 503 KNAELFPEGLRPALEKHSVKSFR-----VYNCTCSSGQVPGAGRPSTPKE 552
 DB 500 R-----FHSQARAKETARQPLDHCVTFTLKCDSKSK--RRGRKSPKXE 544
 QY 553 X-FIVPEFELTNQEVLTASCDLCTIVKTEKRLKALIRLKAHVEQPHQLSGMMD 611
 DB 545 VSHITAEFEITFKVEASDTCADCLRKAEQSLQMA-KTLRKSTGRQCFYQVSGTEVE 604
 QY 612 VAKKPRHTERAEBSGCGVGAENQVSCSAGTGYDCAERELICPNTFGNEGQVNTC 671
 DB 603 VAKKPRHTERAEBSGCGVGAENQVSCSAGTGYDCAERELICPNTFGNEGQVNTC 663
 QY 672 EPCPPRGNSGALKTFEAMNMEGCGGLCCPGEVSADGAPACQCALGTQPEAGRTSCPPC 731
 DB 664 EPCPPRGNSGALKTFEAMNMEGCGGLCCPGEVSADGAPACQCALGTQPEAGRTSCPPC 720
 QY 732 GGGJATKQAGATSPQDCETRVQCSGCHRYNTTHHCICPNVGTYPPEFGKNNVCSPNNT 791
 DB 722 GGGJATKQAGATSPQDCETRVQCSGCHRYNTTHHCICPNVGTYPPEFGKNNVCSPNNT 780
 QY 792 STDPDGSNTITQCKXPRGCEGJGCEFGYIESPNYNGNYPANTECTWTNPPKRRILIV 851
 DB 791 STDPDGSNTITQCKXPRGCEGJGCEFGYIESPNYNGNYPANTECTWTNPPKRRILIV 840
 QY 852 PEIPLPIEDDCGDLVWKRKSSSVTTEYETQTERPIAFTSRKSLMIQKSNEGNSA 911
 DB 841 PEIPLPIEDDCGDLVWKRKSSSVTTEYETQTERPIAFTSRKSLMIQKSNEGNSA 900
 QY 912 RGFQVPPYVTD 922
 DB 901 RGFQVPPYVTD 911

RESULT 15

US-10-221-097-51
 Sequence 51, Application US/10221097
 Publication No. US20030144476A1
 GENERAL INFORMATION:
 APPLICANT: Agarwal, Pankaj
 APPLICANT: Murdock, Paul R.
 APPLICANT: Rizvi, Safia, K.
 APPLICANT: Smith, Randall, F.
 APPLICANT: Xiang, Zhaoying
 TITLE OF INVENTION: NOVEL COMPOUNDS
 FILE REFERENCE: GP50016
 CURRENT APPLICATION NUMBER: US/10/221,097
 CURRENT FILING DATE: 2002-09-26
 PRIOR APPLICATION NUMBER: PCT/US01/07143
 PRIOR FILING DATE: 2001-03-05

PRIOR APPLICATION NUMBER: 60/187,107
PRIOR FILING DATE: 2000-03-06
PRIOR APPLICATION NUMBER: 60/236,874
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/188,916
PRIOR FILING DATE: 2000-03-13
PRIOR APPLICATION NUMBER: 60/237,846
PRIOR FILING DATE: 2000-10-03
NUMBER OF SEQ ID NOS: 52
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 51
LENGTH: 965
TYPE: PRF
ORGANISM: Homo sapiens
US-0-221-097-51

Query Match 58.7% Score 3290.5; DB 12; Length 965;
Best Local Similarity 59.7% Pred. No. 2.5e-241;
Matches 580; Conservative 110; Mismatches 162; Indels 119; Gaps 14;

10 GAA--MAVLLLLLPRLLLAGAVPPGRRAAG-----PQEDVDECAQGLDQCHADAL 61
2 GAAAVRHLVLLAL-----GTRGLAGSGGLPSVADVDECEGDDCHIDAI 49
62 CONTPTSYKSCGKPGYOGEGROCEDIDECGNE-LNGGCVHEDCLNIPGNYRCTCFPGFLA 123
50 CNTPTKSYKLCCKPGYKSGKQCEDIDECENDYNGGCVHACINIPGNYRCTCFDGYLA 109
121 HDGHNCLDVEDECLENNCGCHTCVVMGSEYECCKESFFLSDNCHTCIHRSEGLSCYNK 183
110 HDGHNCLDVEDEQDNGGCGQICVAMKSYECQCHSGFFLSDNCHTCIHRSEGLSCYNK 169
181 DHGSHICKAARPGSVACERPRGFEELANKQSDCITLCHNGSGCGHSCDTADGPSCSH 240
170 DHGCAHICRETPKGGVACDCRPFDAQONKDCITLCHVNGSGCGHSCDEDTDTGPTCCH 229
241 POYKMHDTGRSCLEHEDTVLEVESNTTSVVDGDKVAKRLMETCAVNNGGCDRTCKDT 300
230 QKVALHSDORTCI-----ETCAVNNGGCDRTCKDT 259
361 STGVAGSCPVGFTLLDQKTKCDIDECQTRNGGCDHFCNKTVGSFDGCKKGFKLITDEK 360
260 ATGVAGSCPVGFTLLDQKTKCDIDECQTRNGGCDHFCNKTVGSFDGCKKGFKLITDEK 319
361 SCQDVDECSLRTCHSCINHPGTFACACNMGYTLGFTHCGDNECSINNGGCGVCAV 420
320 TCQDIDECSEFERTCHICINSFSGQLCHRGYLLGFTTHCGDVDECSMSNGSCDQGCYN 379
421 TVGSYECQCHPGYKLMWKKDCVEY-KGLLFCSVPVSLHCGKSGGGGCGFLRC----- 474
380 TKGSYECVCPGPRRLHNMKDCVETGKCLSRKNTSPRAQLSCSKAGVESCFLSCPAHTL 435
475 -----HSGIHLSS--SDVTITRTSVTFKLENGKCSJ 502
440 FVPDSENSYVLSGVPGFGKALOKRNGTSSGLSPSCSDAPTPIKOKARPKIRDAKCHL 499
503 KNAELFPEGRLALPEKSSSVSEFR-----YNNLTSSGKQVPGAPGRPTPKE 552
500 R-----PHSQARAETARQFLDHCHTVFTLKCDSKK--RRGRKSSPSKE 544
553 M-FIVEFELEINQKEVASCGLSCIVRTKRLKAIKTRKAVHREQFHLQSGNNLD 611
545 VSHITAFETIEKHEASDTCEADCLRKRAEOSLCAALKITRKS-GRQFTVQVSGTEYE 604
612 VAKKPRRTSERGASCGVGGQHAENQCVSCRAGTYVYGARERCICPNNGTFQNEEGQMT 672
605 VAORPAKALEGG-ACGAGQVLDQSKVACGPGTHFGSELGQCVSCMPGTYQDWEQJSC 663
672 EPCPRPGSALKTPLEANMNSCGGLQCGEVSADGFAFCQLCALGTPQAPAGRTSCGFC 732
664 TPCP---SSDGLGDPGARNVSECGGCGSPGFFSADGFRPCQACPVGTYQPEPGRGCGFC 720
732 GGGLATKHGATSPQDCETRVQCSPGHFYNTTHRCIACPVGTYQPEPFGKNNCVSCPNT 752

721 GGGLATKHGATSPQDCETRVQCSPGHFYNTTHRCIACPVGTYQPEPFGKNNCVSCPNT 780
792 TPCPQDSTNITQCKXNRCCGELGDFTGYESPNYGNYPANTECTWTINPPKRRLIIV 851
781 STDPODSTNITQCKXNRCCGELGDFTGYESPNYGNYPANTECTWTINPPKRRLIIV 840
852 PEFLIPEDDQGDYLVWRKTSNNSVTTYETQTYERPIAFTSRKSLMIFKSNENSA 911
841 PEFLIPEDDQGDYLVWRKTSNNSVTTYETQTYERPIAFTSRKSLMIFKSNENSA 900
912 RGFQVRYVTVYD 922
901 KGFQVRYVTVYD 911

Search completed: October 22, 2003, 11:40:28
Job time: 97 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 22, 2003, 11:25:58 : Search time 31 Seconds
(without alignments)
1363.502 Million cell updates/sec

Title: US-09-747-371-2
Perfect score: 5605
Sequence: 1 MGAVGRNRPGAAWAVLLLL.....PRFIRLRSVNPSFLRPYX 999

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:
1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCUTOS_COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	692	12.3	956	2	US-08-897-443-3
2	682.5	12.2	638	2	US-08-897-443-1
3	624.5	11.1	1394	6	5177197-30
4	571	10.2	1253	3	US-08-479-722B-4
5	553.5	9.9	1251	5	PCT-US95-02251-3
6	553.5	9.9	1252	1	US-08-199-780-3
7	553.5	9.9	1252	2	US-08-316-650-3
8	552	9.8	1833	3	US-08-479-722B-2
9	552	9.8	1833	5	PCT-US95-02251-18
10	526.5	9.4	2321	4	US-09-230-652-2
11	513.5	9.2	2703	1	US-08-185-432-19
12	513.5	9.2	2703	4	US-08-899-232-4
13	501.5	8.9	2523	4	US-08-185-432-18
14	501.5	8.9	2523	4	US-08-899-232-3
15	491.5	8.8	2556	1	US-08-083-590A-20
16	491.5	8.8	2556	3	US-08-532-384-20
17	485	8.7	2471	1	US-08-185-432-16
18	485	8.7	2471	1	US-08-083-590A-19
19	485	8.7	2471	4	US-08-532-384-19
20	485	8.7	2471	4	US-08-899-232-1
21	483	8.6	2556	1	US-08-185-432-17
22	483	8.6	2556	4	US-08-899-232-2
23	482.5	8.6	1964	4	US-09-467-997-1
24	431	7.7	1055	3	US-09-214-278-2
25	430	7.7	1065	2	US-08-400-159-8
26	429	7.7	1212	3	US-09-214-278-3
27	429	7.7	1257	3	US-08-611-729A-8

28	427	7.6	1238	3	US-09-214-278-5	Sequence 5, Appl
29	426.5	7.6	448	2	US-08-984-072-1	Sequence 1, Appl
30	426.5	7.6	448	2	US-09-212-168-1	Sequence 1, Appl
31	424	7.6	1193	2	US-08-400-159-10	Sequence 10, Appl
32	424	7.6	1193	3	US-08-611-729A-10	Sequence 10, Appl
33	422.5	7.5	1218	2	US-08-400-159-6	Sequence 6, Appl
34	422.5	7.5	1218	3	US-08-611-729A-6	Sequence 6, Appl
35	421.5	7.5	1236	3	US-08-082-046-7	Sequence 7, Appl
36	421.5	7.5	1236	3	US-09-068-740A-6	Sequence 6, Appl
37	421.5	7.5	1187	4	US-09-068-740A-7	Sequence 7, Appl
38	421.5	7.5	1208	4	US-09-199-865-1	Sequence 1, Appl
39	421.5	7.5	1218	4	US-08-882-046-2	Sequence 2, Appl
40	421.5	7.5	1218	4	US-09-068-740A-11	Sequence 11, Appl
41	419.5	7.5	1218	3	US-09-214-278-7	Sequence 7, Appl
42	412	7.4	1219	3	US-08-882-046-5	Sequence 5, Appl
43	412	7.4	3623	2	US-09-141-461-2	Sequence 2, Appl
44	410	7.3	443	2	US-08-833-963C-2	Sequence 2, Appl
45	410	7.3	443	3	US-08-980-514-1	Sequence 1, Appl

ALIGNMENTS

```

RESULT 1
US-08-897-443-3
: Sequence 3, Application US/08897443
: Patent No. 599,263
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Lal, Preeti
APPLICANT: Corley, Neil C.
APPLICANT: Shah, Puri
APPLICANT: Kaser, Mathew
TITLE OR INVENTION: HDVXN MACRILIN-3
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSER: Inocyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/897,443
FILING DATE: Filed Herewith
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy C.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0348 US
TELEPHONE: 415-855-0555
TELECOMMUNICATION INFORMATION:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 956 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 2072792
US-08-897-443-3
Query Match 12.3% Score 692, DB 2, Length 956,
Best Local Similarity 29.8% Pred. No. 1.7e-41.

```

Matches 159; Conservative 72; Mismatches 224; Indels 78; Gaps 14;

QY 62 CONPTSYKSCCKPQY--QGEGRQCEDIDECNELNGCYVHDLNIPANRYCTGDFGML 119
 DB 253 CLNTGSGYICKCKQGYILSTGKCRIDLCATE-DHCEGLCVNMGSPFVQCQYSGYTL 311
 QY 120 AHDGNCIDVDECLNNGCGQHTCVNMGSEYECCKBGFSLSDNCHTCHRSSEGLSCYN 179
 DB 312 AEDGRCTAMDYCASENHGCHECVNMAESSYLRCHEGFALNSKCKTCKID---YCAS 367
 QY 180 KDHGSHCKEAPRGSVACGRPGFELAKNORDC--LTCNHGNGCGQHSODDTADGPEC 237
 DB 368 SNHGQHECVNAD-QTSALCRCLKGFMILPDKTTRINVCALNKGCEHECVNTEBHYC 426
 QY 218 SCHPOYKMTDGRSCJE-----REDTLEVTESNTTSVVDG--DKRYARRLME 284
 DB 427 RCRGGYNLDRPKGKTSRVDHCAQDQHGCEQLCLNTESEFVCCSGEGLINDLKTCSRAC 456
 QY 285 TCAYNNGGCDRTCKTSTGVHSCGFVFTLQDGTCKCJIDECQTRNGGCDHFCANVGS 344
 DB 487 YCLLSNHGCEYSCVNTDKSFACQCFEGHVLRSDDKTAKLDSGALGDHGEHSCVSEDS 546
 QY 345 PDGCKKGFLLTEPKSCQDVDEC-SLDRCTDHCINHPGTFACACRGVTL-YGFTHG 402
 DB 547 FVCCQFEGYILRDGKTRKRDVQDVHNGEHLCVNMGSEYVCKLEGFRLAEDGRCR 606
 QY 403 DTNECSINNCGCQGVCVNTVSGSYECQCHPGYKLMNKKDVE-----444
 DB 607 RKNVCKSTQHGECHNCVNNGNSYLCRCSEGFVLAEDGHCRCRTEGPIDVFFV-DGSKSL 666
 QY 445 -----VKGLFT-SVSPVSLHCGSGSGGDCFLRCHGCHLSSVTLPRISVT 432
 DB 667 GEENFETVKNFVTGLIDSLAVSPKA-----ARVGLLOYSTQVTRTEFT 728
 QY 493 FK---LNEGKSLNKAELPFGELRPALEPKHSSVSKESFRVYVNTCSGKQVP 541
 DB 709 LRGSASAKEMKAVYHMKYMGKSGMTGLALGH-MFERFTQVBGARPRSTQVP 760

RESULT 2
 US-08-897-443-1
 / Sequence 1, Application US/08897443
 / Patent No. 5981263
 / GENERAL INFORMATION:
 / APPLICANT: Hillman, Jennifer J.
 / APPLICANT: Lal, Preeti
 / APPLICANT: Corley, Neil C.
 / APPLICANT: Shah, Purvi
 / APPLICANT: Kaser, Mathew
 / TITLE OF INVENTION: HUMAN MATRILIN-3
 / NUMBER OF SEQUENCES: 4
 / CORRESPONDENCE ADDRESS:
 / ADDRESSEE: Incyte Pharmaceuticals, Inc.
 / STREET: 3174 Porter Drive
 / CITY: Palo Alto
 / STATE: CA
 / COUNTRY: USA
 / ZIP: 94304
 / COMPUTER READABLE FORM:
 / MEDICAL TYPE: Diskette
 / COMPUTER: IBM Compatible
 / OPERATING SYSTEM: DOS
 / SOFTWARE: FASTSEQ for Windows Version 2.3
 / CURRENT APPLICATION DATA:
 / APPLICATION NUMBER: US/08/897,443
 / FILING DATE: Filed Herewith
 / PRIOR APPLICATION DATA:
 / APPLICATION NUMBER:
 / FILING DATE:
 / ATTORNEY/AGENT INFORMATION:
 / NAME: Billings, Lucy C.
 / REGISTRATION NUMBER: 36,749
 / REFERENCE/POCKET NUMBER: PF-0348 US

TELECOMMUNICATION INFORMATION:
 TELEPHONE: 415-855-0555
 TELEFAX: 415-845-4166
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 638 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 IMMEDIATE SOURCE:
 LIBRARY: JTRSN022
 CLONE: 681719
 US-08-897-443-1

Query Match 12.2%; Score 682.5; DB 2; Length 638;
 Best Local Similarity 33.0%; Pred. No. 5e-41;
 Matches 139; Conservative 60; Mismatches 163; Indels 59; Gaps 11;

QY 62 CONPTSYKSCCKPQY--QGEGRQCEDIDECNELNGCYVHDLNIPANRYCTGDFGML 119
 DB 253 CLNTGSGYICKCKQGYILSTGKCRIDLCATE-DHCEGLCVNMGSPFVQCQYSGYTL 311
 QY 120 AHDGNCIDVDECLNNGCGQHTCVNMGSEYECCKBGFSLSDNCHTCHRSSEGLSCYN 179
 DB 312 AEDGRCTAMDYCASENHGCHECVNMAESSYLRCHEGFALNSKCKTCKID---YCAS 367
 QY 180 KDHGSHCKEAPRGSVACGRPGFELAKNORDC--LTCNHGNGCGQHSODDTADGPEC 237
 DB 368 SNHGQHECVNAD-QTSALCRCLKGFMILPDKTTRINVCALNKGCEHECVNTEBHYC 426
 QY 218 SCHPOYKMTDGRSCJE-----REDTLEVTESNTTSVVDG--DKRYARRLME 284
 DB 427 RCRGGYNLDRPKGKTSRVDHCAQDQHGCEQLCLNTESEFVCCSGEGLINDLKTCSRAC 456
 QY 285 TCAYNNGGCDRTCKTSTGVHSCGFVFTLQDGTCKCJIDECQTRNGGCDHFCANVGS 344
 DB 487 YCLLSNHGCEYSCVNTDKSFACQCFEGHVLRSDDKTAKLDSGALGDHGEHSCVSEDS 546
 QY 403 DTNECSINNCGCQGVCVNTVSGSYECQCHPGYKLMNKKDVE-----444
 DB 607 RKNVCKSTQHGECHNCVNNGNSYLCRCSEGFVLAEDGHCRCRTEGPIDVFFV-DGSKSL 666
 QY 445 -----VKGLFT-SVSPVSLHCGSGSGGDCFLRCHGCHLSSVTLPRISVT 432
 DB 667 GEENFETVKNFVTGLIDSLAVSPKA-----ARVGLLOYSTQVTRTEFT 728
 QY 493 FK---LNEGKSLNKAELPFGELRPALEPKHSSVSKESFRVYVNTCSGKQVP 541
 DB 709 LRGSASAKEMKAVYHMKYMGKSGMTGLALGH-MFERFTQVBGARPRSTQVP 760

RESULT 3
 517197-30
 / Patent No. 517197
 / APPLICANT: KANZAKI, TETSUO;LOFSSON, ANDERS;MOREN, ANITA;
 / MERNSTEDT, CHRISTER;HELMAN, ULF;MIYAZONO, KOHEI;CLAESSEN, WELSH,
 / LEVA,HELDIN, CARL;HENRIK
 / TITLE OF INVENTION: ISOLATED NUCLEOTIDE SEQUENCE EXPRESSING
 / HUMAN TRANSFORMING GROWTH FACTOR-BETA1-BINDING PROTEIN
 / NUMBER OF SEQUENCES: 53
 / CURRENT APPLICATION DATA:
 / APPLICATION NUMBER: US/07/487,343
 / FILING DATE: 27-FEB-1993
 / SEQ ID NO:30;
 / LENGTH: 1394
 / 517197-30

Query Match 11.3%; Score 624.5; DB 6; Length 1394;
 Best Local Similarity 24.9%; Pred. No. 1.9e-36;
 Matches 240; Conservative 94; Mismatches 334; Indels 297; Gaps 48;

QY 45 DVECAQGGDDCGIADALCONPTSYKSCCKPQY--QGEGRQCEDIDECNELNGCYVHDLNIPANRYCTGDFGML 119
 DB 253 CLNTGSGYICKCKQGYILSTGKCRIDLCATE-DHCEGLCVNMGSPFVQCQYSGYTL 311
 QY 120 AHDGNCIDVDECLNNGCGQHTCVNMGSEYECCKBGFSLSDNCHTCHRSSEGLSCYN 179
 DB 312 AEDGRCTAMDYCASENHGCHECVNMAESSYLRCHEGFALNSKCKTCKID---YCAS 367
 QY 180 KDHGSHCKEAPRGSVACGRPGFELAKNORDC--LTCNHGNGCGQHSODDTADGPEC 237
 DB 368 SNHGQHECVNAD-QTSALCRCLKGFMILPDKTTRINVCALNKGCEHECVNTEBHYC 426
 QY 218 SCHPOYKMTDGRSCJE-----REDTLEVTESNTTSVVDG--DKRYARRLME 284
 DB 427 RCRGGYNLDRPKGKTSRVDHCAQDQHGCEQLCLNTESEFVCCSGEGLINDLKTCSRAC 456
 QY 285 TCAYNNGGCDRTCKTSTGVHSCGFVFTLQDGTCKCJIDECQTRNGGCDHFCANVGS 344
 DB 487 YCLLSNHGCEYSCVNTDKSFACQCFEGHVLRSDDKTAKLDSGALGDHGEHSCVSEDS 546
 QY 403 DTNECSINNCGCQGVCVNTVSGSYECQCHPGYKLMNKKDVE-----444
 DB 607 RKNVCKSTQHGECHNCVNNGNSYLCRCSEGFVLAEDGHCRCRTEGPIDVFFV-DGSKSL 666
 QY 445 -----VKGLFT-SVSPVSLHCGSGSGGDCFLRCHGCHLSSVTLPRISVT 432
 DB 667 GEENFETVKNFVTGLIDSLAVSPKA-----ARVGLLOYSTQVTRTEFT 728
 QY 493 FK---LNEGKSLNKAELPFGELRPALEPKHSSVSKESFRVYVNTCSGKQVP 541
 DB 709 LRGSASAKEMKAVYHMKYMGKSGMTGLALGH-MFERFTQVBGARPRSTQVP 760

Db 546 EINECTVNPDIIGA-GHCINLPVRYTCIYEGYRSEQKCKVDIDEC-TQVCHLSCGR 603
 QY 102 CLINPGNYRCTCFDGFMLAHGPHNLDEVECEENNGGCGHCVNVMGSEYEC-CKEGRFEL 162
 Db 604 CENTGSGFLC-CPPAGFPASBEGTNCIDVDECBRLPVCGEHCNVTGAFREYCDSGIRM 663
 QY 161 SDNQHTCIHRSEEGSCNWKHGCASHICKKAPRGSVACECPGSELANQPHDCILTNHG 220
 Db 664 TOR-----GRCEIDECINPSTCPBQCVNSP-GSYQC-----VPCREG 701
 QY 221 NGCQHSDDTADGPECSCHPOYKMHDTGRSCLEPEDIIVLEVTESNTTSVVDGDKRVKR 280
 Db 702 FRGMNGQCLDVE-----CLF----- 717
 QY 281 LLMETCAVNNGGCDRTCKDSTGVHSGCPVGFLLDAGTKCDIDECOTRNGGDHPCKN 340
 Db 718 --PNVCA--NG--DCSNLESGTWSCHKGTTRPDHKGCDIDECQGYLQVNGQCKN 769
 QY 341 IVGSPDCGCKKGFLLTDEKSGQDVDECSLDRTCDH-SCINHPGTFAACNRYTLYGF- 398
 Db 770 TEGSFRCTCGGYOLSAKQDCEIDECQHRHLCAGGQCRTEGSGFCVCDQGYRASGLG 829
 QY 399 THGSDTNECSINNCGCQ-VCVNTYVSGTECQCHPEYKILHMK-----KQCVKGLPFS 452
 Db 830 DHCEDINECLEBEXSVCKRGDCINTAGSYDCTCPDGFOLDMKTCCODINEC-EHPGL- 894
 QY 453 VSPRSLHCKGSG--GGDGGF-LRCHSG:HLSSVTTIR-SVTKLNEKCSLKNALF 508
 Db 895 -----CGPGECJNTEGSHVCVCGQGFSTISADGRT-----CDI 918
 QY 509 PEGLRPALPEKH--SSVKEFPYVNLTCSSGKYVP-----GAPGRP- 547
 Db 919 DECVNNTVCDSHGFNDNAGSPRL--CYGQFQFQCGQGVNNEGELLGVCGEAFC 975
 QY 548 SCKEMFITY-----EFLETNKEVTASCDLSCTVATEKPLKKA:RT:RKA----- 595
 Db 976 ENVEGSLFCVACADENQEVSPMTGQCRSRTSLDLYDVQCPKEKKCYVNLNDASLCNV 1035
 QY 596 ----VHREGPHQLS---KMLDVAKKPRPTSERGAESGVCVCGH-----AEN- 636
 Db 1035 LAPNVTKQSCCTSGAKMGWDCNCEIFPCFVLGTAEPTKCPKKGKGVPAEGSSSEAGGENY 1095
 QY 637 ----QCV-----SCRAGTYVDGARFECI-----CQNGTF 662
 Db 1096 KDADECLLFGOEICKNGPCLNTRPGVECYCKQGTYYDPRYKLOCFPMDECCBPSSSIDCQC 1155
 QY 663 QNEGQMTG-----EPCPPRPGNS-----G 681
 Db 1156 VNTGSSVNGFCTHPVWLDASEKRCIRPAESNEQIEETDVYCDLWEHLSDVEVCSRPVWG 1215
 QY 682 ALKT-----PEAMNMECGGGLGQPGEYSADGAFPOQLCALSTFCEAGRTSGF- 729
 Db 1216 KQTTYTECCCLYGEAMGX-QC-ALC-PLKSDSDVQCLNIPVGTGRCP-YGRDALVDPE 1272
 QY 733 ---PCGGGLATKQCATSFQDCETRVQCSFGHFVYTTTHRC:RCFVG-----TYCEP 779
 Db 1272 QYTFEADPYFIODRPLNSFELQAE-EC-G:LNCENGRCVAVDEGTYCDCLDLYHLDT 1328
 QY 780 GKNVCVSCPGNITTTFDGSGTN-ITQCKNRCGGLGDF-----TGYSIS-PNY-PGNY 829
 Db 1329 AKMTCFD-----VNECEDLNRMSLCKNAKCIINTGSGYKCLCLPGVPSDPKPNYCTPANT 1383
 QY 830 PANTJE 834
 Db 1384 A:NLB 1388

RESULT 4
 US-08-479-722B-4
 Sequence 4, Application US/08479722B
 Patent No. 6074840
 GENERAL INFORMATION:
 APPLICANT: Bonadio, Jeffrey

APPLICANT: Yin, Wushan
 TITLE OF INVENTION: LATENT TGF-BINDING PROTEIN (LTBP)
 TITLE OF INVENTION: GENES, COMPOSITIONS AND METHODS
 NUMBER OF SEQUENCES: 13
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Williams, Morgan & Amerson
 STREET: 7676 Hallmont, Suite 250
 CITY: Houston
 STATE: Texas
 COUNTRY: USA
 ZIP: 77040
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 OPERATING SYSTEM: IBM PC compatible
 SOFTWARE: Patent Release #1.0, Version #1.10
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/479,722B
 FILING DATE: 07-JUN-1995
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US PCT/US95/02251
 FILING DATE: 21-FEB-1995
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 09/316,650
 FILING DATE: 30-SEP-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/199,780
 FILING DATE: 18-FEB-1994
 ATTORNEY/AGENT INFORMATION:
 NAME: Fussey, Shelley P.W.
 REGISTRATION NUMBER: 39,458
 REFERENCE/DOCKET NUMBER: 4100.000500/FUS
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (713) 934-7000
 TELEFAX: (713) 934-7011
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1253 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-479-722B-4

Query Match 10.2%; Score 571; DB 3; Length 1253;
 Best Local Similarity 24.5%; Pred. No. 1,1e-12;
 Matches 218; Conservative 83; Mismatches 296; Indels 292; Gaps 46;

QY 32 VPPQGRG---AAGQEDVDECAQGLDDCHADALQNTPTSYKSCKPGYQ3-EGRQCED 86
 Db 539 LPEPSAVELAPQVETDECRNLNONG-CHQCVQPSDYSCHNAGYRSHPCRYCVD 596
 QY 57 IDBCNELNGCYVHCDCNTPGNRCTCFDGFML--AHQHNCLDVEHC-----LENNGCO 140
 Db 597 VNEEAPPCPGKQKICNTGSSYNCHNRRYRUVAGGASCVLNECAKPHLCGGDG- 654
 QY 141 HTCVNVWGSYECCCKEGFFLSDQHTCIHRSEGLSGMNDHGCASHICKAPRGSVAC-S 199
 Db 655 -FCINFGHYKCNCPYRILKASRPICEDIDE---CRDSTGDCGEKKA?-GSFCIA 709
 QY 200 CRPGEAAKQQRDCILTCNHNNGGQCHSCDDTADEPCS---CHPYKMHDTGRSCLER 255
 Db 710 CQPGYR-----SGGGGACR-----DVNECSGTPSPSPG----- 738
 QY 256 EDYVLEVTESNTTSVVDGDKVAKRLLMECAVNNGCDRTCKDSTGVHSGCPVGFLLQ 315
 Db 739 -----CEVLPSSYRCTCAQGRTR 757
 QY 316 LDGKCKDIDECQTRNGSCDHPCKNIVGSPDCGCKKGFLLTDEKSGQDVDECSLDRTG- 374
 Db 758 TGRLSCLDVEDECEGKVCQDQICNTPTGSGTQCCLSYHLSRPSRREDIDBDQFPAACT 817
 QY 375 DHSCLNHPGTFAACNKGTYLYGFTHC-GDTNECSINNCGC-QQVCVNTVGSYECQCHPG 432

```

Db      818 GDCDINTNGSYACJLCPGLRVLGKRRCKKDDIDECSDPGLCLPHACENLQSSYVCVDEG 877
Qy      433 YKLHNKKXDCVFKGLPTSVSPRVLCHGKSGGSDGFLRCHSGITHSSVITRISVT 492
Db      878 FTLTQDHCCEVE-----QPHHKK-----CYLNPDCVFDSD---VLATNVT 919
Qy      493 FTLNKGCSL-----KNAELPPEGJRPALPEKHSSVKEFSFYVNLTTSSGQVPGAPGR 547
Db      919 --QDECCSLGAGMGDHCETIYPCPYSS-AEFHSLVFDGR-----JHSGQO----- 962
Qy      548 SPPKEMFIVEFELETNCKEVTASCDJSCIYKRTKRLKALPLTKAVHNECHLQJLSG 607
Db      963 -----HCEL-CI-----PAHRCIDECILFG 981
Qy      608 XALDVAKRPRTSERQAESCGVCGHAEHCVS-----CRAGTYDGAERPCI----- 655
Db      982 -----AEIC-----KEGKCVNTQPGYECYCKQGFYDGLLCEVDDE 1019
Qy      656 ----LCPNGTFQNEGGQMTCEPCP---RPGNSGALKTPAAMXSE---CGG-CQPG 702
Db      1020 CLDESNCNVCENRGYRCACTPPAEYSPAQCL-IPERWSPGQDVACAGASE--E 1076
Qy      703 VSADGAPQCALGTF-----QPEAGRTSCFPC-----GGGLATKIQGATSPQDC--- 748
Db      1077 RYACVGMWAPAL-TFDGCCCRQRLG-TQCRPCPPRGTSQCTSSSENSTFDTSP 1134
Qy      749 -----ETRVQCSPGHFYNTTTRCI-----RCPVGTIYQPFQKNCVSC 787
Db      1135 LIGKSPRDESDSESDDEC-----RCVSGRCVPRPGAVCECP-GGFQDLASRARCV- 1185
Qy      788 PGNITTPDGSNTITQ---CKNRRCGELGDF-----TGVISS-PNYP 826
Db      1186 -----DIDECRELNRGJLCKSERCVNTSGSFRCVAKGFTRRSPHP 1228

RESULT 5
PCT-US95-02251-3
Sequence 3, Application PC/TUS9502251
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR STIMULATING BONE
TITLE OF INVENTION: CELLS
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESS: Arnold, White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: Texas
COUNTRY: United States of America
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compat.Die
OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII
SOFTWARE: Patent In Release #1.0, Version
SOFTWARE: #1.38
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/02251
FILING DATE: CONCURRENTLY HEREWITH
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/316,650
FILING DATE: 30-SEP-1994
CLASSIFICATION:
APPLICATION NUMBER: US 08/199,780
FILING DATE: 18-FEB-1994
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Parker, David L.
REGISTRATION NUMBER: 32,165
REFERENCE/DOCKET NUMBER: UMIC009P-
TELECOMMUNICATION INFORMATION:

```

```

TELEPHONE: (512) 418-3000
TELEFAX: (713) 799-2679
TELEX: 79-0924
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1251 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US95-02251-3

Query March 9.3%; Score 553.5; DB 5; Length 1251;
Best Local Similarity 24.6%; Pred. No. 21e-31;
Matches 218; Conservative 82; Mismatches 300; Indels 285; Gaps 47;

Qy      32 VPPGRGR---AAGPGEVDCAQZJDDCAGALACONTPTSYKSCSKPYQG--EGROCED 96
Db      537 LPPSRSAVLAETQVETEDCRLNQC-GHQCVPGPSDYSCHNAGYRSHPHRYCYD 595
Qy      27 IEEGNELGCCVCHDCUNIPGNYRCTCEDGFWL--AHQHNCLDVEDC---LENNGGCQ 140
Db      596 VNECEAEPPGPGKGIOMNTGGSYNCHNGRYLHVAGGRSCVDLNECAKPHLCDDG-- 653
Qy      141 HTCVNMGSECCCEGFEJSDNHTCIHSEEGSCNKHGCSHICKEAPRGSYAC-E 199
Db      654 -FCINFGHYKCNCTGYRLKASRPICEDIDE--CRPSTCPGKCNKP-GSFKCA 702
Qy      200 CRPFELAKNRDCLTCHNGGCGHSCDITADGPGSCBQYXNHTDGRSCLREEDTV 259
Db      709 CQGYR-----SQGGACR-----DVNEGS-----EGTFC 733
Qy      260 JEVTSNTTSVVDGKRYRRLMETCAVNGSGCRTCQDSTGVHCSGPVFTLQJCK 319
Db      734 -----SPWCEK---LPGSYRCTAQGIRTRGL 760
Qy      320 TGKDIDECOTRNGGCDHPCKNIYGFDCGCKKGFKLITDEKSCQVDECSLDRTG-DHSC 378
Db      761 SCIDVDDCEAGKVCQGTCTNTPGTFQCCLSGYHLSRPSRCEDIDECPAACTIGDC 820
Qy      379 ZHPTGACACNNGYLYGFTHC-GEINBGS-INNGC-QQCVNTVGSYECCQHPYKXH 436
Db      821 INTNGSYRCPLGRLVGRKCKKDIDCSQDPLCLPHACENLQSSYVCVDEGFTLG 880
Qy      437 MNKDCVEYKGLPTVSPRVSJHCKSKSGGCGCFRCSGJHJSSVTTRESYTFKAN 496
Db      881 QDCHGCEVE-----QPHHKK-----CYLNPCTVFDSD---VLATNVT--QQ 919
Qy      497 EGKCSL-----KNAELPPEGJRPALPEKHSSVKEFSFYVNLTTSSGQVPGAPSTPK 551
Db      920 ECCCISGAGKCHCEIYPCPYSS-AEFHSLVFDGR-----JHSGQO----- 961
Qy      552 EMFIVEFELETNCKEVTASCDJSCIYKRTKRLKALPLTKAVHNECHLQJLSGMLD 611
Db      962 -----HCEL-CI-----PAHRCIDECILFG--- 980
Qy      612 VAKKPRITSERQAESCGVCGHAEHCVS-----CRAGTYDGAERRC----- 655
Db      981 -----AEIC-----KEGKCVNSQPYECYCKQGFYDGLLCEVDDE 1022
Qy      656 -LCPNGTFQNEGGQMTCEPCP---RPGNSGALKTPAAMXSE---CGGLCPGFSYAD 706
Db      1023 SNCRNVGCEHTRALPCACPPAEYSPAQCL-IPERWSPGQDVACAGASE-ERTAC 1076
Qy      707 GFAPQCALGTF-----QPEAGRTSCFPC-----GGGLATKIQGATSPQDC--- 748
Db      1079 VWGPMWAPAL-TFDGCCCRQRLG-TQCRPCPPRGTSQCTSSSENSTFDTSP 1136
Qy      749 -----ETRVQCSPGHFYNTTTRCI---RCPVGTIYQPFQKNCVSCGN-----T 791
Db      1137 SPDESDSESDDEC-----RCVSGRCVPRPGAVCECP-EGGFGQDLASRAR 1191
Qy      792 TFDGSGNTITQ---CKNRRCGELGDF-----TGVISS-PNYP 826

```

Db 1182 CVDIDECRELNGGLCKSERCVNTSGSFRVCVCKAGFTRSRPHGP 1226

RESULT 6

US-08-199-780-3
Sequence 3, Application US/08199780
Patent No. 5763416

GENERAL INFORMATION:

APPLICANT: Bonadio, Jeffrey
APPLICANT: Goldstein, Steven A.
TITLE OF INVENTION: Gene Transfer Into Bone Cells
TITLE OF INVENTION: And Tissues
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: TX
COUNTRY: USA
ZIP: 77210

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release 4.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/199,780
FILING DATE: 18-FEB-1994
CLASSIFICATION: 5.14
ATTORNEY/AGENT INFORMATION:
NAME: Parker, David
REGISTRATION NUMBER: 32,165
REFERENCE/DOCKET NUMBER: UMIC:002
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 320-7220
TELEFAX: (512) 474-7577
INFORMATION FOR SEC. ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1252 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-199-780-3

Query Match 9.9%; Score 553.5; DE 1; Length 1252;
Best Local Similarity 24.6%; Pred. No. 2.1e-31;

Matches 218; Conservative 82; Mismatches 300; Indels 285; Gaps 47;

QY 32 VPRGRGR---AAGPQDVDECAAGLDDCHADALGQKTPSYKSCCKPGYQ--FGQCEPD 86
Db 538 LPPSRSAVEIAPQVETETDECRNLNNTIC-GHGGCVFPGSDVSGCHMAGYRSHPRHYCVS 596
QY 87 IDECGNELNGGCVHDLNIPGNYRCTCFDGFML-AHDGNCIDVDEC---LENNGGCQ 140
Db 597 VNECEAPGCPGKGIOMNTGSGYNCHNNGYRLHVGAGRSCTVDLECKAPHLGCGG-- 654
QY 141 HTCVNMGSYECCKGKGFSLDNQHTCIHSEEGLSCKMKHGCSTCKEAFRGSVAC-E 199
Db 655 -FCINFGHYKCMCYRGLKASRPICEIDE---CRDPSTCPDGCKENKP-GSFKCIA 709
QY 200 CRPGFELAKKQRCILTNHNGGCGHSCDDTDAGPEGCHPYKNTHTGRSLREEDIV 259
Db 710 QDPGR---SOGGACR---DVNECS---EGTFC--- 734
QY 260 LEVTSNTSVVDGKRVKRLIMTCVANNGGCDRTCDTSTGVHSCQPVGFTICDGR 319
Db 735 -----SPGCKE-----LPSYCTCAQGRRTGRL 761
QY 320 TCKDIECCGRNGSCCHFCNNIVGSPDCCGKGFLLTDEKSCQDVDEGSLDRTG-DHSC 378
Db 762 SCIDVDDECEAGKVCQDGICTNTPGSPCCQLSGYH-SRDRSRCEIDEDCDPAACGCGD 821
QY 379 INRGTACACNNGYTLTGFTHC-GDTNCSINNQC-QQVCNVTGVSIECCGHPYKHL 436

Db 822 INTNGSRCLCPGRLHVGGRKCKKIDECGQDPGLCPHACENLGGSVYCVDEGFTLT 881

QY 437 WKKDCEVKGGLPTSVSPVSLHCKSGGGGCGFRCRSHGLHLSQVTTRTSVTFKLN 496

Db 882 CDQHGEVE-----QPHHKE-----CYLNPDDTVFCDS-VLAINVT--00 920

QY 457 EKGCSL-----KNAELFPEGLRALPEKHSVSESRFVYNLTSSGKYVGAAGRSTK 551

Db 921 ECCGSGAGWDHCEVPCPVYSS-AEFSHLPDGR-----LHSGQ----- 962

QY 552 EXFIVEFELETNQKEVTASCDLCTVKTREKALRIPLRKAHVRQGFOLSGMNUD 611

Db 963 -----HCELCI-----PAHRIDECLFG----- 981

QY 612 VAKKPRTSERQAESGCVSGGHAENQVS-----CRAGTYDGRERCI----- 655

Db 982 -----AEIC-----KEGKCVNQPGVEGCKGFPYDGRLECVVDECLDE 1023

QY 656 -LCPTGTFQNEBQKTCPCP-----RPGNGALKTPEAMMSE-----CGGLCPGEXSAD 706

Db 1024 SNRNGVCN-TWRLPACTPRAVSPAQOCL-IPRWSTPQDYKAGASE--ERTAC 1079

QY 707 GFAPQULCAUGT-----QFEAGRTSCFPC-----GGG-ATRGAGATSPDC----- 748

Db 1080 VMGWKAPAL-TRDCCCRQPRUG-TQCRCPKRGTSQCPRTQSSESNSWDTSPLLGX 1137

QY 749 -----TRVQCSFGHRYNTTHRCRCRPGVGTQPEFGKANCYSCPN-----T 791

Db 1138 SPREDSSSEDSERF-----RCVSGPCVPRGAWC-ECPPGFCULASRAR 1182

QY 792 TTFPGGSTNITQ-----CKNRGSGELGCF-----TYIES-PNYP 826

Db 1183 CVDIDECRELNGGLCKSERCVNTSGSFRVCVCKAGFTRSRPHGP 1227

RESULT 7

US-08-316-650-3
Sequence 3, Application US/08316650
Patent No. 5942496

GENERAL INFORMATION:

APPLICANT: Bonadio, Jeffrey
APPLICANT: Roessler, Blake J.
APPLICANT: Goldstein, Steven A.
TITLE OF INVENTION: METHODS AND COMPOSITIONS
TITLE OF INVENTION: FOR STIMULATING BONE CELLS
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: Texas
COUNTRY: USA
ZIP: 77210

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/316,650
FILING DATE: 30-SEP-1994
CLASSIFICATION: 5.14
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/199,780
FILING DATE: 30-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Parker, David L.
REGISTRATION NUMBER: 32,165
REFERENCE/DOCKET NUMBER: UMIC:008
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 418-3600

Db 1344 CG-----FETSPSGNECVDNVECELMMAVCGDALCENVEGSPJC 1382
 QY 588 A:RTLRKAVHREDFHL--QLSGMNLDAVAKKXPRTSERGAES-----CG-- 628
 Db 1383 LCASLEBYDAEBCHCRFRVAG---AQRIPVETLDAQPSLRIMESYSHNGGPPESQ? 1438
 QY 629 VGGHAENOCVSGAGTYDGAERECILCPNGTFQNEEGMTCEPCPR-----PG 673
 Db 1439 LGQNSTGAEC--CCTGGAARMGKA-----CAPGSEDSVEFSQJCS? 1477
 QY 679 NSGALKTPDEAH-----NMSEC---GGICQFGEYSADGFAPCQICAIQTSPAGART 726
 Db 1478 GGGYLPVAGAMTFGGTMYTADAECEVLFGPALCQNGRCS--NIVPGYICL----- 1524
 QY 727 SCFPGGGLATKRGAGTSFQDCETRVGCSGHEFNKTTTHRC-----IRCPVCTYQPE 778
 Db 1525 -----CNPQYHYDASSRKQCCHNECQDLACENGECEVQ? 1557
 QY 779 FGKXNCVSCPGNTTIDPDGNTITQCKARRGCGELGDF-----TGYSIESPR 824
 Db 1558 EGSFHCILCNPPLTDLISGQ---RCVNTTISTE--DFPDHD:HMDCMKKV:NDVCQSP? 1610
 QY 825 YPGNYPANTEC 835
 Db 1611 LRGHHTYTEC 1621

RESULT 10

US-09-230-652-2
 / Sequence 2, Application US/09230652A
 / Patent No. 6537775
 / GENERAL INFORMATION:
 / APPLICANT: Tournier-Lasserre, Elisabeth
 / APPLICANT: Joutelet, Anne
 / APPLICANT: Bousset, Marie-Germaine
 / APPLICANT: Bach, Jean-Francois
 / TITLE OF INVENTION: GENE INVOLVED IN CADASIL, METHOD OF DIAGNOSIS AND
 / TITLE OF INVENTION: THERAPEUTIC APPLICATION
 / FILE REFERENCE: 03115.0048-00000
 / CURRENT APPLICATION NUMBER: US/09/230.652A
 / CURRENT FILING DATE: 1999-05-17
 / EARLIER APPLICATION NUMBER: FR 96 09733
 / EARLIER FILING DATE: 1996-06-01
 / EARLIER APPLICATION NUMBER: FR 97 04680
 / EARLIER FILING DATE: 1997-04-16
 / EARLIER APPLICATION NUMBER: PCT/FR97/01433
 / EARLIER FILING DATE: 1997-07-31
 / NUMBER OF SEQ ID NOS: 163
 / SOFTWARE: Patent in Ver. 2.1
 / SEQ ID NO 2
 / LENGTH: 2121
 / TYPE: PRT
 / ORGANISM: Homo sapiens
 / FEATURE:
 / OTHER INFORMATION: human ADNC No. 6537775ch 3
 US-09-230-652-2

Query Match 9.4%; Score 526.5; DB 4; Length 2121;
 Best Local Similarity 24.2%; Pred. No. 3.8e-29;
 Matches 232; Conservative 75; Mismatches 303; Indels 347; Gaps 58;

QY 33 PFG-RGRAGGCEVDECAOGLDDCHADALCQNPRTSYKSCSRGIVG----- 73
 Db 147 PPGYQGRSC-RSSVDECRVG-EPCRHGGTCLNFGSFRQCPAGYTGPLCEPAVPCAR 203
 QY 80 -----EGRQCE-DIDECGNE--LNGGCVHGCNLIPGNVAC 111
 Db 204 SPGRNGGTCRQSCSLTYDCAALPGEFGNCEVAVDCCPGHRCGLNGGTCVAVGN---TINC 260
 QY 112 TCFGFMALAHGHNCL-DVDECLENNAGCGH--TCVAVGSYECCKEKG--FLSDN-- 163
 Db 261 QCPPEW-----TGCPCTEDVDECCQAPNACHNGGTCFNTLGGHSCVAVNWTGSGSQND 316

QY 164 ---ORTC:R-----SSEGLSCNKHGCSH1CXE-----APRGVA 197
 Db 317 DCAVAFCHGATCHIDRAVASFYCACPMKTKLLHLHDACVSNCHEDA:CDTPVAVGRAI 376
 QY 158 GCPRGGELANQRCDCITLCHNGGCGH--SCDDTADGECSCGHPY--KMHTRGSC 252
 Db 377 CTCPGFGGACDQD-VDECSIGANPEHLGRCVNTGGSFLCCGCRGYTGRCETDNEC 435
 QY 253 J-----EREDYLVETESNTTSSVVDGDRVAVRLLXETGANT---NGCCETTCQDTSVGH 305
 Db 436 LSGCRNATCLDRITGGFTICVAGFTSYCEVDLECGSSPCVNGG---VCDRVNGFS 492
 QY 306 GCPVGFLLQDCKTCK-CEDECCT---RNGGDHCFKNIIVASFCCGCKKGF-LJLDEK 360
 Db 493 CTCFSGFS---STCQLVDDECASTCRGSA---KVVDPDGYECCRAGAFEGCTLCD-- 543
 QY 361 SCQVDDECSLRTCDHS-CINHPGTAFACNRYTLYGFTHC----- 401
 Db 544 --RNVDDSPD-PCMHGRGVVDGIAFSFCACAPGT--GTRCESQVDECRSPCRHGGKC 597
 QY 402 -----GCTN-EGSINNGGCGC-----VCVNTVSGYECQCHGYKLMXKXC 441
 Db 598 LDLYDKYLCRGSSTTGNGCEVNI:DDCASNPCTFGVCRDGINRVDVCQSPF----- 645
 QY 442 CVEVKGJLPTSVSFVSJHCGKSGG--GGGCF:RCHSGIHLSSDVTITRTSVTFKLNKG 498
 Db 650 ---CGPLCNVEINECASFCGEGSGCVDSNGF-RC----- 681
 QY 499 KCSLKNALFPFEGLRP--ALPERKSSVKSFRVYNTLCSGG--KQVPGAGRPSTPKEMF 554
 Db 682 -----LCPGGSJPLCLPSPHCAHE-----PCHGICVDAFAG-- 715
 QY 555 ITVEBELETNKEVTASODLSC:VKRTKRLKAI:RLA:VAHREDFHLQLSGMLDVAK 614
 Db 716 -----FRCV-----CEPGW 724
 QY 615 KPRTSERQAESCGVGGHAENOCVS--CAG--TYVDAREEC:LCPNSTFGNEEGM- 669
 Db 725 SGPRCS-----OSIARDACESQPCRAAGTCSGSDGAFHC-TCPPAV-GARQCELL 772
 QY 670 -TCSPCP-----RPGSGALXTEPMANMSSCGGLQCPGEASAGTAPCC:CALGTF 719
 Db 773 SPCTPMPCHGRCESAPGC:PVSCSPQGMG---GPRCCQVDECCAGPAPCG--PHGIC 926
 QY 720 QPEAGRTSCFPCCGGGLATKHOCA-SFQDCETRVGCSGHEVNTTTHRCIRCFVGYQPEF 779
 Db 927 TNLGSPSC-TCGCG---YTGSPCDDGIN---DQDNPCLNG-----GSCQDGV 868
 QY 780 GKXNCVSCPGNTTIDPDG---STN:TOCKKRRCGGELJGFTGYIES-----PNYRG 927
 Db 869 GSFSQSCJLPG-----FAGPRCARVDECLSNPQPF--GTCTDVASFTCTCPQYGG 918

RESULT 11

US-09-285-412-19
 / Sequence 19, Application US/08185432
 / Patent No. 5750652
 / GENERAL INFORMATION:
 / APPLICANT: Ariavants-Tsakonas, Spyridon
 / APPLICANT: Bousseau, Isabelle
 / APPLICANT: Diederich, Robert J.
 / APPLICANT: Xu, Tian
 / APPLICANT: Matsuno, Kenji
 / TITLE OF INVENTION: DELTA PROTEINS, NUCLEIC ACIDS, AND
 / TITLE OF INVENTION: ANTIBODIES, AND RELATED METHODS AND COMPOSITIONS
 / NUMBER OF SEQUENCES: 23
 / CORRESPONDENCE ADDRESS:
 / ADDRESSEE: PENNIE & EDWARDS
 / STREET: 1155 Avenue of the Americas
 / CITY: New York
 / STATE: New York
 / COUNTRY: U.S.A.
 / ZIP: 10036-2711


```

Db 478 QGTFGRCEYIN:NECESHP---CONEG-SCLDDEGTFFCVMPGFTGCGCEITLID3----- 529
Qy 280 RLMEYCAVNNNGGCDRTCKJTSYGVHSCPVGFLLQDQKTK-DIDECQT---RNSGCC 338
Db 530 -----COSNPELNGTCHKDKINGKSCALGFT-----GACQINIDCCGQPPRRNG-- 577
Qy 336 HFCXIVGSPDCCCKKFKLLTDEKSCQ-DVDECSLDRTCHSHCINHEGT-PACACNRGYT 394
Db 578 -ICHDSIAGVSCCEPPGY-----TGSCCEIN:INDCSNPHRGKIDVNSFKLCIDPGYT 632
Qy 395 LY-----GFTHCQD-----TNESIN--KQCC 424
Db 633 GYICQKQINECESNPOCFDg--HCQDRVGSYYCQCAQTSKNCENVNNECHSPKNNGA 690
Qy 415 QQCVNATVGSYECQCHPGYK-LHMNKKDCVEVKGLLPTSVSPRYSLHCKSGS-----GG 467
Db 691 --TCIDGINSYKCKQCVPGFTGQHCX-----NVDECSPPRANNGVCTIQVNG 736
Qy 468 DGFLECHSGI--HLSSDVTITNTSVTFKLINEGKCS:KNAELF--PEGLR----- 513
Db 737 YKC--ECPRGFYDAHCLSDVDEACASNPC--VNEGRCEGGINFICHCPGYTSKRCLEI 792
Qy 514 ---PALPEKHSSVSESFRVNLITCSSGQVPGAPGRPSTPKEMETVAFLEETNCKE-VT 569
Db 793 DECSNPOCHGCT---CYKLNAPSCQCMPTGYTGO-----KCEINIDCV: 935
Qy 570 ASCDL--SCIWRETEKRLKRAIRLRAKVRHREOFHLQSGNLDVAKP--PRTEROAE 625
Db 836 NPGNGGTCIDK-----VNGYKC-VCKPFTGRCEKSKMD 869
Qy 626 SCGVGQGHAEVOCV-----SCR---AGTYDGARERCIL---CPNG-TFQNEBQ 668
Db 870 POARIRCKEAKCTPSSNFDLPSCCTCKLGYGRYCDDEDIDECSSSPCRNASCINVPGS 929
Qy 669 MTCEFCPR--PENSALKTPEAMNMBEGSGS-CQGEYSADGPAFCUICAGTQPEAGR 725
Db 930 YRC-LCTKGYEGRDAL-----NTDDCASFPQCGRTCCDGI-----GD 967
Qy 726 TSCFPGGGLATKHGATSPQDCETRVQ-----CSPG---HFYNTTHRCICPVQTY 775
Db 968 YSCU-CVDPFGDKH-----CETDINECLSPQNGATSCQVNSYT---CTCPDGL- 1033
Qy 776 QPEFGKNNVCSPGNTTDFDQSTYITQCKARRC--GGEJGP-FTGYIES--PYFENYP 833
Db 1014 --PFGINC-----QTNDEDCITESGCLMGSCIDGINGVSCLAGYSG--- 1054
Qy 831 ANTECTWTIN 840
Db 1055 --ANCCYKLN 1062

```

```

RESULT 13
US-08-185-432-18
Sequence 18, Application US/08185432
Patent No. 5750652
GENERAL INFORMATION:
APPLICANT: Artavanis-Isakoras, Spyridon
APPLICANT: Bussseau, Isabelle
APPLICANT: Diederich, Robert J.
APPLICANT: Xu, Tian
APPLICANT: Matsuno, Kenji
TITLE OF INVENTION: DELTAEX PROTEINS, NUCLEIC ACIDS, AND
TITLE OF INVENTION: ANTIBODIES, AND RELATED METHODS AND COMPOSITIONS
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSES: PENNIE & EDMONDS
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711
COMPUTER READABLE FORM:

```

MEDIUM TYPE: floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/185,432
 FILING DATE: 21-JUN-1994
 CLASSIFICATION: 53C
 AGENT/AGENT INFORMATION:
 NAME: Mistrick, S. Leslie
 REGISTRATION NUMBER: 18,672
 REFERENCE/DOCKET NUMBER: 7326-006
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 869-8664/9741
 TELEFAX: (212) 790-9090
 TELEX: 66141 PENNIE
 INFORMATION FOR SEQ ID NO: 18:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2523 amino acids
 TYPE: amino acid
 TOPOLOGY: unknown
 MOLECULE TYPE: protein
 US-08-185-432-18

Query Match 5.3%; Score 501.5; DB 1; Length 2523;
 Best Local Similarity 23.0%; Pred. No. 2,6e-27;
 Matches 233; Conservative 98; Mismatches 300; Indels 383; Gaps 65;

```

Qy 33 PGRGRARP--QEDVECAQGLDQADALCONTPTSYKSCSKRPYQGEGRCE-DIDE 89
Db 400 PPS---YTGPACNNDVDECSLGANPEHGRGCTNIGSCQCNPGYARP--RCELDVNE 454
Qy 90 CGNELNGGCVHD--CLNIPGNRYCTCFDGFMLAHDSHC-LDVECTLENNSCQHT--CV 144
Db 455 C---LSNPGNDSTCLDQGEFQCIQMPGY---EGLYETNIDECASN--PCLHNGKCI 505
Qy 145 NMGSYECCKKBEFFLSMCHNCIRSEB-----GLSCMKKHCSCSHCKEARG-- 194
Db 506 DKINEFRCDPTGF--SGNLCQHDFTDCTPTCKXNGAKCLDGPSPSYTCCQTEGFTGSH 561
Qy 195 -----SVACGRPGFE-----YAK---NORDE----- 213
Db 562 CEQDINECIPDPHRYGTCDGATFTCLCPREYTGALCENDINECSKCLNGQCTBRE 621
Qy 214 ---ILCNHANGG--CGHSCDDTA-----DGPECSCHPQYKXHTDGRSCLER 255
Db 622 NGYICCPKSTGTGVNCEYK:DDCASLCONKCIKXIDGVECTCEPGY---TGKLC--- 674
Qy 256 EDTVLEVTESNTSVVDQKRYKRLMTCAVNNGGCR-----TCK 298
Db 675 --NININECSNPKRNGG-TCKDQNGFTVCYCPDGYDHMKCLSVNECSNPNCHGACH 730
Qy 299 DSTGVHSCPVGFLLQDQKTKD--DECOFR--NGGCHPCKN:VGSFPGCKKGF 353
Db 731 DGVNGYKDCCEKAGMS---GSNC-D:NNNECESNPKMNGT---CKDYGAYICTGKAG 782
Qy 354 KILTDEKSCQ-DVDECSLDRTCHS-----CYNHF-- 392
Db 783 ---SGPNQTNINECSNPKLHSGTCLDVAQYKCKMLPYTGA:CEAVLAPCAQSPCK 838
Qy 383 -----GTFACACNRGYTLVFTGCGTNECS:VNGGCGQ--VGVNATVGSYECQ 429
Db 839 NGRCKESEDFFETFSCEPPGM--QQTCEIDNBEV--NRPCKNATCQNTNGSYKNC 894
Qy 430 HPGYKJHNNKDC-VEVKGLLPTSVSPRYSLHCKSGSGGCGFLRCHSGIHLSDDVTIR 498
Db 895 KPGY---TSRNEEMIDDCQGNP-----CHNG----- 918
Qy 489 TSVTFKLINEGKCS:KNAELF--PEGLRPALPEKHSVSESRVYNLTQSSGQY--- 540
Db 919 ---GSCS-DGIMNFFCNCPAGFRG--PKCELDINEC---ASNPKCKNANTDCVY 963
Qy 541 -----POARIRPSTPKEMETVAFLEETNCKE:TAACQJSC:YKTEKRLKRAIRTL 592

```

```

Db 964 NSYTCQCGFGSG-----IHESNTPDCTES---SCNGST----- 996
Qy 593 RKAHREOPHLOLQSGNLDVAKKPRTRSERQASGCVGGGAENOCYS--CRAFTYYDGA 650
Db 997 -----CIDGINFTCCQCPGFTGSGYCO-----HDINECSKPLNG----- 1032
Qy 651 RERCILCPNGTFONEGQMTCEPCRPONSGALKTPKAW--NMSEC--GGLOQPEYSADG 707
Db 1033 -----GTQCDSTGYTKC--TCPOGTGLNQCULVRWCDSPPCKNGKC-----KCTIN 1278
Qy 708 FAPQQLCALG--TFQPEAGRTSCFPGGGLATKHQATSPQDCETRVQCSGPHFYNTTH 765
Db 1079 FYRCE--CKSGMTGVCYCDVPSVSC-----EVAKKQGVVDIYHLCRNSGMCVD---TGNTH 1128
Qy 766 RCIRCPVGTIPEFGKNCVSCP-----GNTTFPDG-----STNITQC 804
Db 1129 FC-RCQAG-YTGSYCEQVDECSFNPQNGATCTDYLGSYCECVAGYGVNCSSEINEC 1186
Qy 805 KMRRC--GGELGDFGTGYIESPNYPKVPANT-----ECT-----WTINP 841
Db 1187 LSHPCQNGSTCIDJLN-----TYKSCFRGTGCHCEINVDCTPFYDSFTLFP 1235

```

```

RESULT 14
US-08-899-232-3
Sequence 3, Application US/08899232
Patent No. 6436650
GENERAL INFORMATION:
APPLICANT: Artavanis-Tsakonas, Spyridon
APPLICANT: Qi, Huijin
TITLE OF INVENTION: ACTIVATED FORMS OF NOTCH AND METHODS BASED THEREON
FILE REFERENCE: 7326-046
CURRENT APPLICATION NUMBER: US/08/899,232
CURRENT FILING DATE: 1997-07-23
NUMBER OF SEQ ID NOS: 4
SOFTWARE: Patent Ver. 2.0
SEQ ID NO 3
LENGTH: 2523
TYPE: RRT
ORGANISM: Xeropus sp.
US-08-899-232-3

```

```

Query Match 8.9%; Score 501.5; DB 4; Length 2523;
Best Local Similarity 23.0%; Pred. No. 2,6e-27;
Matches 223; Conservative 98; Mismatches 300; Indels 389; Gaps 65;

```

```

Qy 33 PPGGRRAAP--QEDVDECAQGLDCHADALCQNTPTSYKSCCKQYQSGRQCE--DIDE 99
Db 400 PPG---YTGPAQNNVDDECSLQANPCHEGRCNTLGSFQCNCPQYAGP--RCEIDVNE 454
Qy 90 CGNELNGGCVHD--CLINPGNYRCTCFGFMALAHGHC--LDVDECLENGGCGHT--CY 144
Db 455 C--LSNPQNDSTCLDQIGFQPCICMPY---RGLYETIYIDCAHN--PCLHNGKCI 505
Qy 145 NYMGSYECCKEKFLLSDNQHTCIHRSE-----GASCKMKHGHSHLCKEAPR3-- 194
Db 506 DKINERCDQPTGF---SGNLQGHDFDCTSTPCCKGAKCLDGNSTYCCQCTEFTFRH 561
Qy 195 -----SVACGCRPGE-----LAK--NQKDC----- 213
Db 562 CQEDINECI PDPCHYGTCKDG:ATFTCLCRPGYGTGLCNDINECLSKCLNGGQCTRE 621
Qy 214 ---ILTCNKGNG--CGHSCDPTA-----DQPECSCHPOYMHGHDGSCLER 255
Db 622 NGYICTCPKGTQVNCSTKIDDCASNLCKNGKCIKIDYECTCTEPRY---TGKLC--- 674
Qy 256 EDTVLEVTESNTSYVDGDKRYRR-LMETCAVNNGGCDR-----TCX 298
Db 675 ---NININCDSPKCNNG--TCXQDQINGFTVCVCPDGYHDMCLSEVNECNSPCHGACH 730
Qy 299 DNSTGVHSCSPVFTLQJGKCKKD--DECCTR---NGGCHPCKN:VGSDDCGCKXGF 353

```

```

Db 711 DGWYKCDCEAGWS-----GSNC--DINNCEESNPQNMGT---CKMXTGAVICTCKAGF 782
Qy 354 KILTBKSCQ--DVDECSLDR:CDHS-----CINHP-- 382
Db 783 ---SGPNCQTNINECSNPCLNHGTCIDVAGYCNCKM:PTTGAIICAVLAPCAGSCK 838
Qy 383 -----GTFCACNRYTLVFTHCDGTNECSINMGCCQ--VCVNTVSGYEQC 429
Db 839 NGGRCKESDPEFSECPGWM--QQQTEIIMNECV--NRPCRNAGATCQNTNGSYKNC 894
Qy 430 HPGYKLMWKKDC--VEWKL:PTSVSPVSLHCKSGGGGCGFLACHSHI:HSVTTIR 488
Db 895 KPGY---TGRCNEMDIDCQPNP-----CHNG----- 918
Qy 489 TSVTFKLNKSKSLKALFP-----PEG:RALP:PKHSSVYKESFRVNLTCSSQKV--- 540
Db 919 ---GSCS--DQIMFPNCPAGFRG--PKBEDINEC--ASNPCKNAGNCTDVC 963
Qy 541 ---PGAPGRPTPKXEPITVEFELETNQKVTASCD:SCIVKTEKRLKRAITL 592
Db 964 NSYTCQCGFGSG-----IHESNTPDCTES---SCFNGT----- 996
Qy 593 RKAHREOPH:CLSNLDVAKKPRTRSERQASGCVGGGAENOCYS--CRAFTYYDGA 650
Db 997 -----CIDGINFTCCQCPGFTGSGYCO-----HDINECSKPLNG----- 1032
Qy 651 RERCILCPNGTFONEGQMTCEPCRPONSGALKTPKAW--NMSEC--GGLOQPEYSADG 707
Db 1033 -----GTQCDSTGYTKC--TCPOGTGLNQCULVRWCDSPPCKNGKC-----KCTIN 1078
Qy 708 FAPQQLCALG--TFQPEAGRTSCFPGGGLATKHQATSPQDCETRVQCSGPHFYNTTH 765
Db 1079 FYRCE--CKSGMTGVCYCDVPSVSC-----EVAKKQGVVDIYHLCRNSGMCVD---TGNTH 1128
Qy 766 RCIRCPVGTIPEFGKNCVSCP-----GNTTFPDG-----STNITQC 804
Db 1129 FC-RCQAG-YTGSYCEQVDECSFNPQNGATCTDYLGSYCECVAGYGVNCSSEINEC 1186
Qy 805 KMRRC--GGELGDFGTGYIESPNYPKVPANT-----ECT-----WTINP 841
Db 1187 LSHPCQNGSTCIDJLN-----TYKSCFRGTGCHCEINVDCTPFYDSFTLFP 1235

```

```

RESULT 15
US-08-033-590A-20
Sequence 20, Application US/03003590A
Patent No. 5786158
GENERAL INFORMATION:
APPLICANT: Artavanis-Tsakonas, S. et al.
TITLE OF INVENTION: Therapeutic And Diagnostic Methods
TITLE OF INVENTION: And Compositions Based On No. 5786159ach Proteins And
NUMBER OF SEQUENCES: 2;
CORRESPONDENCE ADDRESS:
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/033,590A
FILING DATE: 25-JUN-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Miarock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7326-015

```

TELECOMMUNICATION INFORMATION:
 TELEPHONE: 212 780-9090
 TELEFAX: 212 8698864/9741
 TELE: 66141 PENNIE
 INFORMATION FOR SEQ ID NO: 20:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2556 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: unknown
 MOLECULE TYPE: peptide
 OS-08-093-590A-20

Query Match 8.8%, Score 491.5; DB 1; Length 2556;
 Best Local Similarity 23.7%; Pred. No. 1.4e-26;
 Matches 238; Conservative 90; Mismatches 339; Indels 337; Gaps 57;

```

QY 41 GP--QEDVDECAAGLDGCHADALCNTPTSYKSGCK----- 74
Db 172 GPTCRQDVNECGQKRLCRHGGTCHNEVGSIRVCVRATHTSPNCEPRYPVPCSPSCQNG 231
QY 75 -----PSYQGEGRQCEIDIEC-GNEL-NGSCVHDCLN-PGNYRCTCFDGF 117
Db 232 TCRPTGVTHECACLPFTQNGE-ENIDCPGNCKNGACVGVN---TYNCPCEPW 287
QY 118 MLAHDGHNCL-DVDECLENNCGQH--TCVVMGSEYECCKEGFF---LSDN-----QHT 165
Db 288 ----TGQYCTEDVDECOLMPNACONGTCHNTGGYNCVCVNGWTGEDCSENIDCASAA 343
QY 167 CIH-----RSEGLSCMKNDHGCSECKEAP-----RGSVACECRPG 203
Db 344 CFHGATCHDRVASFCEPHGRTGLCHLNDACISNPNESNCDTNPVNGKALCTCPSC 403
QY 204 FELAKNRPDCI-TCNHNNGGCGHS--CDDTADGPECSCHPY---KMTDGRSCL---E 254
Db 404 YTGPAQSGD-VDECSLGANPCEHAGKICNTLGSFEQCCLQGYTGPRCEIDVNECVSNPQC 462
QY 255 REDTYLEVTESENTSVVDGDRVKRRLMETCA----VNNGCDRTCKDTSTGVHSCPV 310
Db 463 NDATCLDQIGFQCMCPGYEGHCEVNTDECASSPCLHNG----RCLDKINEFQCECFT 518
QY 311 GFTLQDGTCK-DIDEQCT---RNGGCDHFCNK-VGSEFDCGKKGFKLLTDEKSCQD 366
Db 519 GFT---GHLQYDVDECASTPCKNGA---KCLDGPNTYTCVCTBGY---TGTHCEVDID 568
QY 367 ECSLDRTCDHSCINPRTFAACNKGTYLVGFTHC-GDTNECSIN---NGCQCG----- 416
Db 569 ECDPDPCHYGSCKDVATFTCLCRGYTGH---HCEFNINECSSQPCSLRGTCQDPNAY 625
QY 417 -----VCNNTVSGSECOCHPGYKLMKKKDCVEVKGL 448
Db 626 LCFCLKTGGPNCNELNDCCASSPCDSGTCDKTDGYEACEPGTSMONSNIDECAG- 684
QY 449 LPTSVPRVSLHCGKSG---GGDGCFLRCHSGIHLSSDVTITRTSVTFKLINEKCSLXN 524
Db 685 ----NF-----CHMGTCEDGINSFTCRCEGYH---DPTCLS-----EYNE--CN--- 721
QY 505 AELPPEGLRPALEPHYSSVKESFRVNLTCSSGKOVFGAPGRPSTPKEMFTVBELELTN 564
Db 722 -----SNFCVHAGACRDLNSYKCDG-----PCMSG----- 747
QY 565 QKEVTASCDLS-----CIYKRTFKRLKAKIRTLKRAVHRECFHJOLSGMNLVAKKP 616
Db 748 -----TNCIDINNENESNPCVNGGTCKDMTSGI---VCTGREGF---SGPN----- 787
QY 617 FRTSRCAESGCVGQGHAEHQVSCRAGTYYD--GAREFCILCENGTFQNEEGOMTCE- 672
Db 788 ---CQTNINEC-----ASNPCIN--KGTCLIDVAGYKNCCLPYTGA-----TCBY 826
QY 673 ---PC-PRP-GNSGALKTPBAMNMSGC-----GSLCQPSYSADGFAPCCLCALGTF 715
Db 829 VLAFCAPSPFCRNGSECESEYSEFSVCPTAGAKQTCCEV-DINECVLSPCRHGA---- 883

```

```

QY 726 QPEAKRSCFPCGGG-ATKHQATSPDCCEPRV-QCSPGHFY-----NTTHRCIR 769
Db 884 -----SCQNTGGYRCHQAGYSGRNCETDIDCRBNPCHNGSGSTDDGINATACDCLP 936
QY 770 CPVGYQPEFGKKNQVSCP---GNTTID-----FDG---STNITQCKNRRC-- 809
Db 937 GFRGTFCBE-D-NECASDPGRNAGNCTDCVDSYCTCTPAGTSGIHCENNTPTCTESSCFN 995
QY 310 GSE---LGDFTGYIESPNYPANYPANTCTWTINPPRRRLI 849
Db 996 GGTVCVQINSFT-CLCPGFTSSY-----CQHVAVECDSEKPCCLD 1033

```

Search completed: October 22, 2003, 12:31:49
 Cdb time : 39 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OK protein - protein search, using sw model

Run on: October 22, 2003, 11:24:13 ; Search time 121 Seconds

(without alignments)
2130.533 Million cell updates/sec

Title:

US-09-747-371-2

Perfect score:

5635

Sequence:

1 MGAVGRNRGAAMVLLLL.....PRSF-RLRSKSRFLRPYK 999

Scoring table:

BLOSUM62

Searched:

Gapex 10.0 , Gapext 0.5

Total number of hits satisfying chosen parameters:

830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_23:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp Vertebrate:*
- 14: sp Unclassified:*
- 15: sp_virus:*
- 16: sp_bacteria:*
- 17: sp_archaea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5635	100.0	999	4 Q9N36	Q9N36 homo sapien
2	5056	90.2	997	11 Q9J30	Q9J30 mus musculu
3	3662.5	65.3	988	4 Q81W4	Q81W4 homo sapien
4	3640.5	65.0	992	11 Q8C9Q4	Q8C9Q4 mus musculu
5	3397	60.6	993	4 Q81X30	Q81X30 homo sapien
6	3262.5	58.2	961	11 Q9EQC6	Q9EQC6 mus musculu
7	2916	52.0	880	4 Q8NAU9	Q8NAU9 homo sapien
8	1738.5	31.0	581	4 Q8NAU8	Q8NAU8 homo sapien
9	875	15.6	1664	5 Q9TVQ2	Q9TVQ2 caenorhabdi
10	841.5	15.0	1574	11 Q88281	Q88281 ratuS nov
11	775.5	13.8	1469	5 Q9V589	Q9V589 drosophila
12	737.5	13.2	1246	4 Q75095	Q75095 homo sapien
13	720	12.8	2906	11 Q9WU9	Q9WU9 ratuS nov
14	710.5	12.7	2909	4 Q96TP8	Q96TP8 homo sapien
15	695	12.4	810	11 Q8C435	Q8C435 mus musculu
16	695	12.4	956	11 Q99K64	Q99K64 mus musculu

17	695	12.4	956	11 Q8R542	Q8R542 mus musculu
18	695	12.4	1289	5 Q8SS53	Q8SS53 dictyosteli
19	694	12.4	2189	5 Q9B105	Q9B105 eimeria ten
20	693	12.4	3857	11 Q88840	Q88840 mus musculu
21	686.5	12.2	937	4 Q96FT5	Q96FT5 homo sapien
22	686.5	12.2	1016	4 Q8NDE6	Q8NDE6 homo sapien
23	675	12.0	2872	11 Q9WU98	Q9WU98 ratuS nov
24	654.5	11.7	61	4 Q9Y559	Q9Y559 homo sapien
25	640	11.4	708	13 P87363	P87363 galluS gall
26	610.5	10.9	1062	11 Q60789	Q60789 mus musculu
27	610	10.9	1398	13 Q8AXM6	Q8AXM6 xenopus lae
28	607	10.8	1399	13 Q8JF24	Q8JF24 xenopus lae
29	601	10.7	1389	11 Q8CG18	Q8CG18 mus musculu
30	601	10.7	1713	11 Q8CG19	Q8CG19 mus musculu
31	598	10.7	1713	11 Q88349	Q88349 mus musculu
32	597.5	10.7	1963	6 Q28319	Q28319 bos taurus
33	587	10.5	1382	4 Q9H7K2	Q9H7K2 homo sapien
34	587	10.5	1972	5 Q8WPL1	Q8WPL1 oikopleura
35	583.5	10.4	1764	11 Q35806	Q35806 ratuS nov
36	578	10.3	293	6 Q8WJK0	Q8WJK0 cercopithe
37	575.5	10.3	733	5 Q25678	Q25678 podocoryne
38	575	10.3	1551	5 Q9NGV4	Q9NGV4 drosophila
39	575	10.3	3396	5 Q9VM55	Q9VM55 drosophila
40	574	10.2	1174	11 Q95K58	Q95K58 mus musculu
41	571	10.2	1253	11 Q61810	Q61810 mus musculu
42	570.5	10.2	1821	4 Q14767	Q14767 homo sapien
43	567	10.1	658	4 Q8NBH6	Q8NBH6 homo sapien
44	567	10.1	685	11 Q922K8	Q922K8 mus musculu
45	566.5	10.1	746	4 Q96HB9	Q96HB9 homo sapien

ALIGNMENTS

RESULT :

ID Q9N36 PRELIMINARY: PRT: 999 AA.

AC Q9N36; 31-OCT-2000 (TREMBLrel: 15, Created: 31-OCT-2000 (TREMBLrel: 15, Last sequence update: 31-MAR-2003 (TREMBLrel: 23, Last annotation update: CEGP1 protein.

GN CEGP1.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

OX NCBI_taxid=9606;

RN [1]

RP SEQUENCE FROM N.A.

RA Bahr A., Hankeln T., Winterpacht A., Zabel B., Schmidt E.R.; Comparative sequencing of Human chromosome 1p15 and mouse chromosome 7.

RT Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.

CC -1- SIMILARITY: CONTAINS 1 CUB DOMAIN.

DR EXBL; A040877; CAB92285.1; -.

DR HSSP; P07204; IADX.

DR InterPro: IPR000152; Asx hydroxyl.

DR InterPro: IPR000659; CUB domain.

DR InterPro: IPR001881; EGF_CA.

DR InterPro: IPR006209; EGF-like.

DR Pfam; PF00431; CUB; 1.

DR Pfam; PF00006; EGF; 9.

DR SMART; SM00342; CUB; 1.

DR SMART; SM00179; EGF_CA; 6.

DR PROSITE; PS0010; ASX_HYDROXYL; 6.

DR PROSITE; PS01180; CUB; 1.

DR PROSITE; PS01186; EGF_2; 8.

DR PROSITE; PS01187; EGF_CA; 6.

KA EGF-like domain.

SC SEQUENCE 999 AA; 109956 MW; 61314844A053095 CRC64;

Query Match 100.0%; Score 5635; DB 4; Length 999;

Best Local Similarity 100.0%; Pred. No. 0;

Matches	999	Conservative	0	Mismatches	0	Indels	0	Gaps	0
QY	1	MGVAGNRPGAAAVLLLLLLP.LLLAGAVPGRGRAGPQEDVDECAQGLDDCHADA	60						
DB	1	MGVARNRPGAAAVLLLLLLP.LLLAGAVPGRGRAGPQEDVDECAQGLDDCHADA	60						
QY	61	LCQNTPTSYKSCSKPGYQGEGRQCEIDIDECGNELNGGCVHDCLNIPGNYRCTCFDGFMLA	120						
DB	61	LCQNTPTSYKSCSKPGYQGEGRQCEIDIDECGNELNGGCVHDCLNIPGNYRCTCFDGFMLA	120						
QY	121	HDGHNCLDVDECLENNGGCHTCVNMVMSYECCKEGFF.SDNQHTCHIRSEEGLSGMNK	190						
DB	121	HDGHNCLDVDECLENNGGCHTCVNMVMSYECCKEGFF.SDNQHTCHIRSEEGLSGMNK	190						
QY	181	DHGGSHICKKAPRGVACCECRPGFELANQDCLITCNHNGNGGCHSCDDTADGPECSC	240						
DB	181	DHGGSHICKKAPRGVACCECRPGFELANQDCLITCNHNGNGGCHSCDDTADGPECSC	240						
QY	241	POYKMTDGRSCLEREDTVLEVTESNTTSVVDGDKRVKRLJMETCAVNNGGCDRTCKDT	300						
DB	241	POYKMTDGRSCLEREDTVLEVTESNTTSVVDGDKRVKRLJMETCAVNNGGCDRTCKDT	300						
QY	301	STGVHSCSPVGFLLDLDGKTCKDIDECOTRNGGCHPCKNTVGSFDDGCKKGFLLTDEK	360						
DB	301	STGVHSCSPVGFLLDLDGKTCKDIDECOTRNGGCHPCKNTVGSFDDGCKKGFLLTDEK	360						
QY	361	SCQDVDECSLDRTCDSHCINHPGTACACNRGYTLVGFTHCGDGTNECSINNGGCQYCVN	420						
DB	361	SCQDVDECSLDRTCDSHCINHPGTACACNRGYTLVGFTHCGDGTNECSINNGGCQYCVN	420						
QY	421	TVGSYECQCHPGYKLMNKKDCVEYKGLPTSVSPVSLHCGSGGGGGLRCHSGIHL	480						
DB	421	TVGSYECQCHPGYKLMNKKDCVEYKGLPTSVSPVSLHCGSGGGGGLRCHSGIHL	480						
QY	481	SSDVTTIRTSVTFKLENGKCSLKNALFPEGRLPALPEKHSVKESPRYVNLTCSSGKQV	540						
DB	481	SSDVTTIRTSVTFKLENGKCSLKNALFPEGRLPALPEKHSVKESPRYVNLTCSSGKQV	540						
QY	541	PGAPGRSTPKKEMFTVFELETNQEVYASCDLSICVKTREKRLKKAIRTLRAVHREQ	600						
DB	541	PGAPGRSTPKKEMFTVFELETNQEVYASCDLSICVKTREKRLKKAIRTLRAVHREQ	600						
QY	601	FHLQSGMNLVAKKPPRTSEPOASCGVGGHANQCVSCRAGTYGYABERCLQNG	660						
DB	601	FHLQSGMNLVAKKPPRTSEPOASCGVGGHANQCVSCRAGTYGYABERCLQNG	660						
QY	661	FHLQSGMNLVAKKPPRTSEPOASCGVGGHANQCVSCRAGTYGYABERCLQNG	660						
DB	661	FHLQSGMNLVAKKPPRTSEPOASCGVGGHANQCVSCRAGTYGYABERCLQNG	660						
QY	721	PEAGRTSCFPCCGGLATKQCATSFQDCETRVQSGFHFVNTTHRCIRACVGTYPBFG	780						
DB	721	PEAGRTSCFPCCGGLATKQCATSFQDCETRVQSGFHFVNTTHRCIRACVGTYPBFG	780						
QY	781	KXNVSCPGNTTDDGSGTNTQCKNRGCGELGFTGYIESPRYPGNYANTETWTIN	840						
DB	781	KXNVSCPGNTTDDGSGTNTQCKNRGCGELGFTGYIESPRYPGNYANTETWTIN	840						
QY	841	PPKRRILIVPEILFIEDDCGDTLWKRKTSSSVTTTECOYEPPIAFTSSKKLM	900						
DB	841	PPKRRILIVPEILFIEDDCGDTLWKRKTSSSVTTTECOYEPPIAFTSSKKLM	900						
QY	901	LOKSNENGSARGFVYVYDDEYQELIEDIVRDGRLYASENHQELKDKKLKALEDV	960						
DB	901	LOKSNENGSARGFVYVYDDEYQELIEDIVRDGRLYASENHQELKDKKLKALEDV	960						
QY	961	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
DB	961	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
QY	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
DB	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
QY	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
DB	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
QY	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
DB	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
QY	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
DB	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
QY	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
DB	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
QY	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
DB	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
QY	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
DB	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
QY	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
DB	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
QY	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
DB	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
QY	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
DB	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
QY	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
DB	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
QY	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
DB	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
QY	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
DB	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
QY	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
DB	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
QY	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
DB	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
QY	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
DB	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
QY	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
DB	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
QY	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
DB	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
QY	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
DB	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
QY	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
DB	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
QY	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
DB	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
QY	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
DB	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
QY	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
DB	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
QY	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
DB	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
QY	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
DB	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
QY	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
DB	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
QY	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
DB	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
QY	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
DB	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
QY	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
DB	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
QY	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
DB	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
QY	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
DB	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
QY	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
DB	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
QY	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
DB	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
QY	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
DB	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
QY	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
DB	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
QY	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
DB	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
QY	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
DB	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
QY	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
DB	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
QY	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
DB	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
QY	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
DB	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
QY	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
DB	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
QY	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
DB	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
QY	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
DB	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
QY	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
DB	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
QY	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
DB	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
QY	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
DB	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
QY	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
DB	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
QY	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
DB	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
QY	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
DB	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
QY	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
DB	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
QY	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
DB	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
QY	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
DB	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
QY	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
DB	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
QY	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
DB	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
QY	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
DB	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
QY	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
DB	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
QY	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
DB	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
QY	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
DB	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRPYK	999						
QY	999	LAHPONYFKYTAQESREMPBSFLLLSKYSRFLRP							

```

Db      SSDVTVRTSVTFKINECKSLQAKLSPEGRLPALPRHSSVSEFOYANTLTSFGKOV 538
Qy      541 PGAPGRPTPKEMFTTFELETNOKEVTASCDLSICVKE-TEKR.ARAIRLRAVAREQ 600
Db      539 PGLALRLNAPKEMFTTFELETNOKEVTASCDLSICVKE-TEKR.ARAIRLRAVAREQ 598
Qy      601 FHLQSGMLDVAKKPRPTSEROAESCQVGGHAEENOCVSCRACTYYDGAERECILCPNG 660
Db      599 FHLQSGMDLMAKTPRSVSGOHEETGCGVGHESQCVSCRACTYYDGAERECILCPNG 658
Qy      661 TFQNEBQMTCEPCRPCNSGALKTPAMNMSDEGGLCPGEYSADGAPQOLCALSTFC 720
Db      659 TFQNEBQMTCEPCRPCNPENLGLSKISEAMNVSDCGGLCPGEYSANGAPQOLCALSTFC 718
Qy      721 PEAGRTSCFPGGGGATGAGATSPQDCETRVQSPGFYNTTHRCRCRCVGTGCEFG 730
Db      719 PDVGRISCLSGGGGLPTKHLGATSPQDCETRVQSPGFYNTTHRCRCRCVGTGCEFG 778
Qy      781 KNCVSCPGNTTTFDGSNTITCKNRRCGELDFTGY-ESPNYPGNYPANTECTWTIN 840
Db      779 KNCVSCPGNTTTFDGSNTITCKNRRCGELDFTGY-ESPNYPGNYPANTECTWTIN 838
Qy      841 PPKRRLIIVPELFLIEDCGDYLVRKTSNSVTVTCTCTYERPIAFTRSKLW 900
Db      839 PPKRRLIIVPELFLIEDCGDYLVRKTSNSVTVTCTCTYERPIAFTRSKLW 898
Qy      901 IOPKSNEGNSARGFQVPTVYDEDEYQELIEDIVRQGLYASENHQELIKKKLKALFDV 960
Db      899 IOPKSNEGNSARGFQVPTVYDEDEYQELIEDIVRQGLYASENHQELIKKKLKALFDV 958
Qy      961 LAHPQNYFYTAGESREMPFRSFRILRLSKVSRFLRPYK 999
Db      959 LAHPQNYFYTAGESREMPFRSFRILRLSKVSRFLRPYK 997

RESULT 3
Q81WY4 PRELIMINARY; PRT; 988 AA.
AC Q81WY4:
DT 01-MAR-2003 (TREMblrel. 23, Created)
DT 01-MAR-2003 (TREMblrel. 23, Last sequence update)
DE Signal peptide-CUB-EGF-like domain containing protein 1.
GN SCUBEL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_taxid=9606;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=2236413; PubMed=12270931;
RA Yang R.B., Ng C.K.D., Maeserum S.M., Colman S.D., Shenoy S.,
RA Mehrahan F., Komuves L.G., Tomlinson J.S., Topper J.N.;
RT "Identification of a Novel Family of Cell-surface Proteins Expressed
RT in Human Vascular Endothelium."
RL J. Bio. Chem. 277:46364-46373(2002).
DR EMBL; AF525689; AAN77133.1; -.
SQ SEQUENCE 988 AA; 107901 MW; 1BF57BEF786C9C1 CRC64;

Query Match 65.3%; Score 3662.5; DB 4; Length 988;
Best Local Similarity 62.2%; Pred. No. 2,9e-317;
Matches 652; Conservative 133; Mismatches 164; Indels 119; Gaps 14;

Qy :0 GAA---MAVLLTLLPLLLAGAVPPGRGAAG-----PQEDVDECAQGLDQCHVAL 61
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 2 GAAVVRHLCVLLAL-----GTRGLAGGSGLPQSVVDVDSBQTDQCHVAL 49
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 62 CNPTSYKSCCKPYQGEGRQCEIDECGNE-LNGSVHCLNIPGVYRGTCTGQWYLA 120
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 50 CQNAKSYKCLCKPQKGGKQCDIDECNENYNGGCVHECINPGVYRGTCTGQWYLA 109
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 12 HDGHNCLDVDCLENNGGQJHTCVNVKSGSYECCKEGFPLSDNQHTCHRSSEBGLSKNK 160

```

```

Db      110 HDGHNCLDVDCLENNGGQJHTCVNVKSGSYECCKEGFPLSDNQHTCHRSSEBGLSKNK 169
Qy      181 DHGSGIICKEARSGVACECRPGFELAKNORDCITCNHNGNGCQHS-CDPTADGEGSC 240
Db      170 DHGSAHICRETFKGVACDCRPGFPLAQNDQCTITCNVNGGCGHSEDDITGTCGGH 229
Qy      241 PQYKHTDGRSCLEREDTVLEVTESNTTSVVDGKRVKRLMETCAVNNGCDBRTCKDT 300
Db      230 QYVAPHSDBRTCL-----ETCAVNNGCDBRTCKDT 259
Qy      301 STGVHSCGVGFTLLQDGKTCQDIBQCTRNCGCHFCNKIVGSPDCCGKAFKULTDEK 360
Db      260 ATGVRCSCVGFLLQPDGKTCQDINECVNNGCCHFCANTVGSEFCGCRKQYKULTDER 319
Qy      361 SCQDVDEGLDPTCHSCINHGTFACACNRYTLXGFHCDTJEGCSINNGCQGVN 420
Db      320 TQDDIDEGSFEKTCQICINSPGSPQCLCHROYILGTHGCGVDECSKNSGDCQGVN 379
Qy      421 TVGSEYECQHPQYKLMNKKDCVEY-KGLLPTSVSPRYSLHGKSGGSGDFLRC----- 474
Db      380 TKSYSQVCPQGRRLHMNKDCVEFGKCLSRKAKTSPPAQLSCKAGVESFCLSPAHFL 439
Qy      475 -----HSGTHLS---SDVTTIRTSVTFKINECKSL 502
Db      440 FVPDSNSYVLSCGVPGQKALQKRNQTSGLGSCSDAPRTPIKQXAFKIRAKCHL 499
Qy      503 KNAELFPESCLRALCEKESVYKESR-----YNNLCSSSGKQVAGARGPTPK 552
Db      500 R-----FHSQRAKETARPQLHCHVTFVTLKCDSSKK--RRGRKSPSK 544
Qy      553 V-FIVEPELEFETNOKEVTASCDLSICVKE-TEKR.ARAIRLRAVAREQ 600
Db      545 VSHITHEFIEFTKEMASDQCEADCCRKRAEQLGALITLAKSIORQGFYGVSGTEVE 604
Qy      612 VAKKPRPTSERQASQGVGCGHAEENOCVSCRACTYYDGAERECILCPNG 900
Db      605 VAKPRPAKLEGGG-ACGAGQVQLQDSKCVACGPTFGGELGQVCCVPQTYDMGQQLSC 663
Qy      672 EPCPRRNGSGLKTPKAMKNSGCGGLCPGEYSADGAPQOLCALSTFC 960
Db      664 TPCP---SSDGLGLPQARVNSDEGCGSGSFGFSADGAPQOLCALSTFC 920
Qy      732 GCGJATKAGATSPQDCETRVQSPGFYNTTHRCIRCPVGTGYPEFSKNKNCVSCPGNT 791
Db      722 GCGJLTGHEGTTSPQDCETRVQSPGFYNTTHRCIRCPVGTGYPEFSKNKNCVSCPGNT 780
Qy      792 TDFDGSNTITCKNRRCGELDFTGY-ESPNYPGNYPANTECTWTIN 840
Db      782 TDFDGSNTITCKNRRCGELDFTGY-ESPNYPGNYPANTECTWTIN 838
Qy      852 PEFLPIEDCGDYLVRKTSNSVTVTCTCTYERPIAFTRSKLW 900
Db      841 PEFLPIEDCGDYLVRKTSNSVTVTCTCTYERPIAFTRSKLW 898
Qy      912 RGFQVPTVYDEDEYQELIEDIVRQGLYASENHQELIKKKLKALFDVLAHPNYFYKT 972
Db      901 RGFQVPTVYDEDEYQELIEDIVRQGLYASENHQELIKKKLKALFDVLAHPNYFYKT 960
Qy      972 ADESSREMPFRSFRILRLSKVSRFLRPYK 999
Db      961 ADESSREMPFRSFRILRLSKVSRFLRPYK 998

RESULT 4
Q8C9Q4 PRELIMINARY; PRT; 992 AA.
AC Q8C9Q4:
DT 01-MAR-2003 (TREMblrel. 23, Created)
DT 01-MAR-2003 (TREMblrel. 23, Last sequence update)
DE Signal peptide.
GN Kux musculinus (Youse).
OS

```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathia; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N. A.
 RC STRAIN=C57BL/6J; TISSUE=Thymus;
 RX MEDLINE=22354683; PubMed=12466851;
 RA The FANTOM Consortium.
 RA The RIKEN Genome Expedition Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RL Nature 420:563-573(2002)."
 SR EMBL; AK041580; BAC00996.1;
 SQ SEQUENCE 992 AA; 108570 XM; 42219D667979.TA1 CRC64;

Query Match 65.0%; Score 3640.5; DB 1.; Length 992;
 Best Local Similarity 63.3%; Pred. No. 2,763,35;
 Matches 658; Conservative 103; Mismatches 180; Indels 99; Gaps 15;

QY 13 GAA---MAVJLLLLPPLLALAGAVPPGRGAAAPQEDVDECAQGLEDCHADALCCATP 66
 DB 2 GAAAVRMHLYLLLLALGRGLVSGSLPG-----AVDDECEGTDCHIDALCQNTP 54
 QY 67 TSYKSCKPGYQSGRQCEIDIECGNE-LNGCVHDCANI PGNYRCTCFDGFMLAHGHN 125
 DB 55 KSYKCLCKPGYKSGRQCEIDIECGNDYNGCVHDCINIPGNYRCTCFDGFMLAHGHN 114
 QY 126 CLDVDECLENNNGGQCHTCVNVVSGYECCEKEGFLSDNQHCHIRSEGLSCMKKHCQS 185
 DB 115 CLDVDECLENNNGGQCHTCVNVVSGYECCEKEGFLSDNQHCHIRSEGLSCMKKHCQS 174
 QY 186 HICKEAPRGSVACRPRFELAKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQK 245
 DB 175 HICKEAPRGSVACRPRFELAKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQK 234
 QY 246 HTDGRSCLEFEDTLEVTESNTSVVDQKRVKRRLLYETCAVNVNGGCDTCKDTSTGVH 305
 DB 235 HADGRITC-----ETCAVNVNGGCDTCKDTSTGVH 264
 QY 306 CSCPVGFTLQDGTCKDIDECQTRNGGCHFCNKIVSGFDQCGKGFLLTDEKSCQY 355
 DB 265 CSCPVGFTLQDGTCKDIDECQTRNGGCHFCNKIVSGFDQCGKGFLLTDEKSCQY 324
 QY 366 DECSLDRTCDHSCINHPCTFACNRCGTYLGFTHCGDTHNCS-INNGCCQVCCNTYGSY 425
 DB 325 DECSLDRTCDHSCINHPCTFACNRCGTYLGFTHCGDTHNCS-INNGCCQVCCNTYGSY 384
 QY 426 ECGCHPGYKLMNKKDCVYKGLPTS-VSPRVSJHCGSGGCGCFLRG----- 474
 DB 385 ECGCHPGYKLMNKKDCVYKGLPTS-VSPRVSJHCGSGGCGCFLRG----- 444
 QY 475 HSGIHLSSDV-----TTTISTVTFNLNKGCS--KKA 525
 DB 445 ESSYILSCGVPGGLGKTLPRKNNGSSSTGPGSCDAPETPIAKARFPIPAKCHILQRSQ 504
 QY 506 ELFPREGIR-PAPEKHSSVKESFRVYVLTGSSGQVPGAPGPRSTPKVY-FITTEPLET 563
 DB 505 ERAKDTCHAPLIDNCHVT-----FVTLKQSSSKT-RRRGRKSPSEVSHITAEFEVEM 556
 QY 564 NQKEVTASCDLSCTIVKRTKRLRAIRTLRAVHREGFHQLQSGNLDVAKKPERTSERQ 623
 DB 557 KYDEASGTCEADCMKRAEQLQAIKILRKISGRNGYVGLTEYEVAGRPKALAGT 616
 QY 624 AESGCGGSHAEHNGCVSRAGTYIDGAREKILCPNGTQNEEQQVCCPPRRGNSGA 683
 DB 617 G-TGGIGQIQDGGKVPAPGTYSGDGGQCMPCVSGTYQVEEQQLSCTPP--SSSECL 672
 QY 684 KTPKAMNSECGLQCPGSEYASGAPCOLCALGTFPEAGRTSCFCGGGLATKRGAT 743
 DB 673 GLAGARNVSECQCGSPGYFSADGKFPQACPVGTIQPEPRGTGFCGGLTKHGT 732
 QY 744 SPQDCETRVGSPGFVYVTTTRICIRCPVGTYQPEFGKNNVSCPGATTTDFDSTNITQ 803

DB 733 SPQDCETRVGSPGFVYVTTTRICIRCPVGTYQPEFGKNNVSCPGATTTDFDSTNITQ 792
 QY 804 CNRRQCELOCPFYIESPNYPNYPANTCTWTINPPKGRILLIVPEFLPIEDCG 863
 DB 793 CKHQCGCELOCPFYIESPNYPNYPANTCTWTINPPKGRILLIVPEFLPIEDCG 852
 QY 864 DYLVRKSSSSNSVYVTCCTYERPIAFTRSKMLTQFNSNGNSARQGVYVYTD 922
 DB 853 DYLVRKSSSPVYVTCCTYERPIAFTRSKMLTQFNSNGNSARQGVYVYTD 912
 QY 923 ---EDYCELEIDVDRGLVASENHCETLKQKLLKALFCVLARPKYFKYTAQSEME 979
 DB 913 RUPEDYQULIEVDSGLVASENHCETLKQKLLKALFCVLARPKYFKYTAQSEME 972
 QY 980 PRSFIRLRSKVSRLRPRYK 999
 DB 973 PRSFIRLRSKVSRLRPRYK 992

RESCUT 5
 DB 08IX30 PRELIMINARY; PRT; 993 AA.
 AC 08IX30;
 DT 01-MAR-2003 (Trembl, 23, Created)
 DT 01-MAR-2003 (Trembl, 23, Last sequence update)
 DT 01-MAR-2003 (Trembl, 23, Last annotation update)
 DE CUB and EGF containing protein.
 GN CEGF3.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N. A.
 RA Pfarr N., Bahr A., Cichutek A., Joebbert R., Zabel B. U., Schmidt E. R.,
 RA Hackel T., Winterpacht A.;
 RA "Novel human gene family (CEGF) encoding mosaic proteins with EGF-
 RT like, STR2 and a CUB module: cloning and expression analysis."
 RL Submitted (NOV-2002) to the EMBL/GenBank/DBS databases.
 DR EMBL; AF452494; AAT6828.1;
 SQ SEQUENCE 993 AA; 109282 MW; 19BB0E85627EAF4 CRC64;

Query Match 60.6%; Score 3397; DB 4; Length 993;
 Best Local Similarity 58.3%; Pred. No. 4e-233;
 Matches 596; Conservative 136; Mismatches 212; Indels 78; Gaps 9;

QY 22 LPPLLALAGAVPPGRGAAAPQEDVDECAQGLDCHADALCQNTPTSYKSCKPGYQSEG 81
 DB 6 VPGCLIVLTVARAQAQSKAAQVDECEGTDNCHIDA-CQNTPTSYKICKSGYTDG 65
 QY 92 RQCEIDIECGNE-LNGCVHDCANI PGNYRCTCFDGFMLAHGHNLDVDECLENNNGGCH 141
 DB 66 KACQVDDEGERDNGCVHDCVNI PGNYRCTCFDGFMLAHGHNLDVDECLENNNGGCH 125
 QY 142 TCNVVSGYECCEKEGFLSDNQHCHIRSEGLSCMKKHCQSIIKEAPRGVACER 201
 DB 126 TCNVVSGYECCEKEGFLSDNQHCHIRSEGLSCMKKHCQSIIKEAPRGVACER 185
 QY 202 PGFELAKQ 261
 DB 186 PGFELAKQK 237
 QY 262 VTESNTSVYSGDKKVRRLMETCAVNVNGGCDTCKDTSTGVHSGGCVGFTLQDGTCK 321
 DB 238 -----ETCAVNVNGGCDTCKDTSTGVHSGGCVGFTLQDGTCK 275
 QY 322 KQIDECQTRNGGCHFCNKIVSGFDQCGKGFLLTDEKSCQVDECSLDRTCDHSCIN 381
 DB 276 KQIDECQTRNGGCHFCNKIVSGFDQCGKGFLLTDEKSCQVDECSLDRTCDHSCIN 335
 QY 382 PGTACACNRGTYLGFTHCGDTHNCS-INNGCCQVCCNTYGSYECQCHPGY-KLHMKK 440
 DB 336 PGTACACNRGTYLGFTHCGDTHNCS-INNGCCQVCCNTYGSYECQCHPGY-KLHMKK 395

QY 441 DCVEVKGLPTSVSPVSLHCGKSGGDCGFLRC----- 474
 DB 396 DTEPLKCGSGPAGKAMLSNRSKDKDTCLTCRRAFLPESENGFTVSGTPSPRAA 455
 QY 475 -----HSGIHLSSD-----VTTIRTSVTFKLNESKC--SLKNAELFPEGRLPALPERHS 521
 DB 456 PARAGHNGNSTSNHCHAEAVLSIKORASFRIKDKACRLHLRNKSKTEBAGRITSPGAP 515
 QY 522 SVKESFRVNLTCSSGKQVPGAPRPRSTPKENF-ITVEFEJTNKXETATASCDLSCTIKR 586
 DB 516 CSECQVTFHLKCDSSRRKGRARRPTPEKVEVTRLLLEHAEVRAEETTAACGLPCCJQR 575
 QY 581 TEKRLAKAIRLTKAHREROFHQLSGMNLJVAKKPPTSEBOA---ECCGVQGHAEHQ 617
 DB 576 MERRLKSLKLRKINSINDRFLRLAGLDYELAHKPLGLVAGRAAPMSSCPGQRAAGT 635
 QY 638 CVSCRAQTYDGAERCC-LCPNGTQNEBGMTCBPCEPNCNGALKTPEAMNMSGGL 697
 DB 636 CVSCPGTYHGGTBOCVCPAGTFOEREGQLSCDLCEGSDAHGFL---GATNVTTCAQC 692
 QY 698 COPGEVSADGFAPCOLAGTFCPAGRTSCFPCGGG-LATKQGATSFQDDETRVQSG 757
 DB 693 CPGQSVGVCFKPCPCPRGTYQPPAGRTLCFPCGGG-LTKHEGALISFQDCDTKYQCSFG 752
 QY 758 HEYNTTHRCIRCPVGTQPEFGKNNCVSCPNTTDDPGSTN-TQCKNRSGGELGFT 817
 DB 753 HYNTSTHRCICMGSTYOPDFRQNPSCSCPRAN-STYDQSTSVACQCRCCJSLGFT 812
 QY 819 GYIESPNYPGNYPANTECTWTINPPKRLILVVEIF-LPEDCGDYLVWKRTSSSNV 877
 DB 813 GYIESPNYGNYPAGCECIWNTNPPKRLILVVEIF-LPSDECGDYLVWKXSSPS 972
 QY 878 TVYEQCTYERPIATSRKSLM-IQKSNESGSAFGVPTTYTTEYQELIEDIVRCGR 937
 DB 873 TVYEQCTYERPIATSRKSLM-INFKSEANSAGFOF-PVYTYCEYQELIEDIVRCGR 932
 QY 938 LYASENHOELIKDKLILKFLVLAHPONYFYTQESREVPASIFILMSKVSRLFR 957
 DB 933 LYASENHOELIKDKLILKFLVLAHPONYFYTQESREVPASIFILMSKVSRLFR 952
 QY 998 YK 999
 DB 992 YK 993

RESULT 6
 Q9EOC6 PRELIMINARY; PRT; 961 AA.
 AC Q9EOC6:
 DT 01-MAR-2001 (TREMblrel. 16, Created)
 DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)
 DE 01-MAR-2003 (TREMblrel. 23, Last annotation update)
 DE EGF-related protein SCUB1.
 GN SCUB1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_taxid=10090;
 RN NCBI_taxid=10090;
 RP SEQUENCE FROM N.A.
 RC STRAIN=3H1;
 RX MEDLINE=20541712; PubMed=11087664;
 RA Grimmond S., Lazder R., Van Hateren N., Siggers P., Hulsebos T.J.M.,
 RA Arkell R., Greenfield A.,
 RT "Cloning, Mapping, and Expression Analysis of a Gene Encoding a Novel
 RT Mammalian EGF-Related Protein (SCUB1).",
 RL Genomics 73:74-81(2000).
 CC -!- SIMILARITY: CONTAINS 1 CUB DOMAIN.
 DR EMBL: AF276425; AKG25939.1; -.
 DR HSSP: P00742; IHG.
 DR MGD: MGI:1890616; Scubel.
 DR InterPro: IPR000152; Asx_hydroxyl.

DR InterPro: IPR000859; CUB domain.
 DR InterPro: IPR001881; EGF Ca.
 DR InterPro: IPR006209; EGF like.
 DR InterPro: IPR001491; Thrombomodulin.
 DR Pfam: PF00431; CUB; 1.
 DR Pfam: PF00008; EGF; 9.
 DR PRINTS: PR00507; THROMBOMODULN.
 DR SMART: SMC0042; CUB; 1.
 DR SMART: SMC0179; EGF CA; 6.
 DR PROSITE: PS00010; ASX_HYDROXYL; 6.
 DR PROSITE: PS01180; CUB; 1.
 DR PROSITE: PS01186; EGF_2; 6.
 DR PROSITE: PS01187; EGF_CA; 6.
 DR ECF-like domain.
 KW SEQUENCE 961 AA; 103989 MW; 966726C4E3BBD33E8 CRC64;
 Query Match 58.2%; Score 3262.5; DB 11; Length 961;
 Best Local Similarity 60.9%; Freq. No. 1.3e-281;
 Matches 584; Conservative 99; Mismatches 181; Indels 95; Gaps 14;
 QY 10 GAA---WAVLLLLLPRLLAGAVPRGRRAGPQEEVCEAGSLDCHADALQNTP 66
 DB 2 GAALVRHLSLLALGARQGVVSGSLPG-----AVDVECSBGTDCHDALQNTP 54
 QY 67 TSYKCSKPGYOGGRQCEIDECQNE-LNGSCVHCLNIPGNVRCYCFDGFYLAHDGN 125
 DB 55 KSYKCLCKPGYKSGRCEIDECQDYNGGVHDCIN-FGVNCTCFDGFYLAHDGN 114
 QY 126 CLVDDECLNNGSCQHTCVNVMGSECCKEGFFLDNCHTCHIREEGSLCNMKDHGS 185
 DB 115 CLVDDECLNNGSCQHTCVNVMGSECCKEGFFLDNCHTCHIREEGSLCNMKDHGS 174
 QY 186 HICKEARSSVACEPRGFEFLANQRDCLTCYHNGSCGHSQDPTADGEGSCHPQYM 245
 DB 175 HICRETPKGVACDPRGFLAQNQDCLTCYHNGSCGHSQDPTADGEGSCHPQYM 234
 QY 246 HTDRSCLEREETVLEVTESNTSVVDGKRVKRLMETCAVNGSGDRCTDSTGVH 305
 DB 235 HADGRCTI-----ETCAVNGSGDRCTDSTGVH 264
 QY 306 CSCPVGFTLQLDGKTCIDCEQTRNGGCHFKNIYGSFDGCKKGFYLLDEKSCQDV 365
 DB 265 CSCPVGFTLQLDGKTCIDCEQTRNGGCHFKNIYGSFDGCKKGFYLLDEKSCQDV 324
 QY 366 DEGSJERTCHSCINHGTPACACNRYTLKSTTHGCDTNECSINNGSCQVCNNTGVS 425
 DB 325 DEGSJERTCHSCINSGSFQCLCRGYTLVGTTHGCDVDECSMNGSCQVCNNTGVS 384
 QY 426 ECGCHPGYKJHMKKDCVEVKGLPTS-VSPVSLHCGKSGGDCGFLRC----- 474
 DB 385 ECVCPGRRLHNNQKDEVNMGCLSRKSAQAQOLSCGKGVENGLSLGHSLFMPDS 444
 QY 475 HSGIHLSSDY-----TTIRTSVTFKLNESKSL-KNA 505
 DB 445 ESSYILSGGVPLQGLKTLPRNGTSSSTGPGCSDAPTPPIRKARFIKIRAKCHLQPSQ 504
 QY 506 ELPFEGGR-PALPERKSSVSESFRVYVNLTCSSGKQVPGAPRPRSTPKEM-FITVEFELET 563
 DB 505 ERANDTLRAPLQNCVY-----FVTLKQDSKSK-RRGRKSPSKVEHITAEFEVEM 556
 QY 564 NQKVTASCCVSCIVKTEKRLKRAITRLKAVHREOFHQLSGMNLJVAKKPPTSERQ 623
 DB 557 KVDBASGTCENQCKRKAQESLOAATITLKSIGRNOFYQVLTREVAGRPKALEGT 616
 QY 624 AESCGVQGHAEHQCVSCRAQTYDGAERCC-LCPNGTQNEBGMTCBPCEPNCNGAL 683
 DB 617 G-TCGIGC-LQDGKCVPCABGTAFSGDPGCGCVSGSTVQDMESQJLCTPCP---SSSG 672
 QY 684 KTPRANMVSFCGSLQCGEYSADGFAFQCALCTPQPEAGRTSCFPCGGGGLATKHGAT 743
 DB 673 GLAGARVSSCGQCGSGVFSADGFAFQCALCTPQPEAGRTSCFPCGGGGLATKHGAT 732
 QY 744 SFQDCEFRVQCSRQHFYTTTHRCIRCPVGTQPEFGKNNCVSCPNTTDDPGSTNITQ 803

Db	733	SPDCEAKNACSGHHYNTTTRCIRCPGTCGPEFGNHCISCPRNTS:DFDGSNTVTH	792
OY	804	CKNRRCGELGDTGYIESPNPGNPANTECTWTINPPKRRLIVPELPIEDDGC	863
Db	793	CKNHCCEGELGDTGYIESPNPGNPANTECTWTINPPKRRLIVPELPIEDDGC	852
OY	864	DYLMRRTSSNVTEYTCYRERATSSKKLMTQFSNGBSARGCVPYUYN	922
Db	853	DYLMRRTSSNVTEYTCYRERATSSKKLMTQFSNGBSARGCVPYUYN	911
RESULT 7			
OANAT9			
12	Q8NAU9	PRELIMINARY;	PRT: 880 AA.
AC	Q8NAU9;		
DT	01-OCT-2002 (TrEMBLrel. 22, Created)		
DT	01-OCT-2002 (TrEMBLrel. 22, Last sequence update)		
DT	01-MAR-2003 (TrEMBLrel. 23, Last annotation update)		
DE	Hypothetical protein FLJ34743.		
OS	Homo sapiens (Human)		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
CC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
CC	NCBI_Taxid:9606;		
PN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Ninomiya K., Magatsuma M., Kanda K., Kondo H., Yokoi T., Kodaira H.,		
RA	Fuuya T., Takahashi M., Kikawa E., Omura Y., Abe K., Kamihara K.,		
RA	Katsuda N., Sato K., Tanikawa M., Yamazaki M., Sugiyama T., Ito R.,		
RA	Csaki T., Sato H., Wakamatsu A., Ishii S., Yamamoto C., Isori Y.,		
RA	Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,		
RA	Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Murakawa K.,		
RA	Kanehori K., Takahashi-Fujii A., Ohtsuna A., Sugiyama A., Kawakami B.,		
RA	Suzuki Y., Sugao S., Nagahara K., Masuno Y., Nagai K., Isogai T.;		
RT	"NEDO human cDNA sequencing project."		
RL	Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.		
CC	-1 SIMILARITY: CONTAINS 1 CUB DOMAIN.		
CC	EXBL: AK092062; BAC03798...		
DR	InterPro: IPR000157; Asx_hydroxy1.		
DR	InterPro: IPR000859; CUB_domain.		
DR	InterPro: IPR001881; EGF_Ca.		
DR	InterPro: IPR006239; EGF-like.		
DR	InterPro: IPR006210; EGF.		
DR	InterPro: IPR004391; Thrombomodulin.		
DR	Pfam: PF00431; CUB: 1.		
DR	Pfam: PF00008; EGF: 5.		
DR	PRINTS: PR00907; THROMBODULIN.		
DR	SMART: SMC0042; CUB: 1.		
DR	SMART: SMC0181; EGF: 5.		
DR	SMART: SMC0179; EGF_Ca: 5.		
DR	PROSITE: PS00010; ASX_HYDROXYL: 3.		
DR	PROSITE: PS01180; CUB: 1.		
DR	PROSITE: PS01186; EGF_2: 5.		
DR	PROSITE: PS01187; EGF_Ca: 3.		
DR	Hypothetical protein.		
DR	SEQUENCE 880 AA; 97239 MW; FQBDDEE85266FD60 CRC64;		
Query March 52.0%; Score 2916; DB 4; Length 880;			
Best local similarity 57.44%; Pval.No. 9e-25;			
Matches 518; Conservative 127; Mismatches 186; Indels 72; Gaps			
OY	146	VMGSECCCKEKEGFLSNDRTCIRSEEDLSCKNCH3CSHICKELAPRGVACGCGFE	2050
Db	1	VMGSECCCKEKEGFLSNDRTCIRSEEDLSCKNCH3CSHICKELAPRGVACGCGFE	60
OY	206	LAKNRCCLTCNHNCGGCHSCDDTADGPEGSCHPDYKARTDGRSCLEBCTVLEVTS	2655
Db	61	LTKNRCCLTCNHNCGGCHSCDDTADGPEGSCHPDYKARTDGRSCLEBCTVLEVTS	1284
OY	266	NTSIVAGDGRVVRRL-----LMEICANNNGGDRCTCMTDTSGVNCSGPAVFTLLDQKT	322
Db	109	-----GERRLECHLPITQAVSNSETCAVANNGGDSKCHGDAAGVCHCTPVAFFMLQDFRKT	161

ID	GENAB6	PRELIM_NAME	PRG	581 AA
AC	GENAB6			
AC	01-OCT-2002 (TREMBlrel. 22, Created)			
DC	01-OCT-2002 (TREMBlrel. 22, Last sequence update)			
DC	01-MAR-2003 (TREMBlrel. 23, Last annotation update)			
DE	Hypothetical protein FLJ14691.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
CC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
CX	NCBI_TaxID=9606;			
RN	111			
RP	SEQUENCE FROM N.A.			
RA	Tashiro H., Yamazaki W., Watarabe K., Kumagai A., Itakura S.,			
RA	Fukuzumi Y., Fujimori Y., Komiyama M., Sugiyama T., Irie R.,			
RA	Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,			
RA	Kawai H., Hiro Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,			
RA	Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Nagasuna M.,			
RA	Mutakawa K., Kanehori K., Takahashi-Niujii A., Oshima A., Sugiyama A.,			

RA Kawakami B., Suzuki Y., Sugano S., Nagahara K., Masuho Y., Nagai K.,
 RA Isogai T.;
 RT "NEO human cDNA sequencing project";
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: CONTAINS 1 CUB DOMAIN.
 DR EMBL: AK092010; BAC03789.1; -;
 DR InterPro: IPR000859; CUB_domain.
 DR Pfam: PF00431; CUB; 1.
 DR SMART: SM00042; CUB; 1.
 DR PROSITE: PS01180; CUB; 1.
 DR Hypothetical protein.
 KM SEQUENCE 581 AA; 64343 MW; D179C28CDA1E4BE1 CRC64;
 SQ
 Query Match 31.0%; Score 1738.5; DB 4; Length 581;
 Best Local Similarity 55.7%; Pred. No. 4,8e-146;
 Matches 325; Conservative 94; Mismatches 129; Indels 47; Gaps 7;
 459 LHCQSGGGDGCFLRC-----HSGHLSDD--- 483
 Db 2 LSCNRGKKCTCALTCPARFLPSENGFTVSCPTSPRAPARAGHNGNTSNHCE 61
 Qy 484 --VTTRTVTFRKLHGKC--SLKNAELFPGGRPALPKHSSVSEFRVNLTCSSGK 539
 Db 62 AAVLS-KORASLKIDAKCRHLNRKGTKEAGRTGPGADQCEQVTFHLKCDSSRK 121
 Qy 540 VPGAPGFSTPKKEMF-ITVEFELETQKQVLTASCDLSCIVKTEKRLKATITLRAVHR 598
 Db 122 GKGRRARTPGKEVTRLTLELNAVPAEETLASCGLRGEMERLKSLMLKSLNS 181
 Qy 539 EGFHLQLSGMNLDAKKPRPTSEQA--ECCGVGGGAENOCVCRAGTYVDGARERC 655
 Db 182 DFFLLRLAGLDYEALHKFGLVAGERAPYECRPGHRAGTQVSCPGQTYHGGTEQGV 241
 Qy 656 LCPNGTFQNEEGGCTCEPFRPNQNSGALKTPRANMSSECGALCQREVSADDFACQDCA 715
 Db 242 PCPAGTFOEREGQLSDCLCPGDADGRL--GATVTTTACGCPPOGSHVDGFKSCQPTP 298
 Qy 716 LGTFOPEAGRTSCFPCCGGGLATKQCATSFQDCETFPVOCSPGHFVTTTHRCISCPVGT 775
 Db 299 RGTQPEARLTICFPCCGGGLTKHGAISFQDDTKVQCSPGHYNTSIHRCIRCMGSI 358
 Qy 776 QPEFGANCVSCPGNTTTPDGSNTITQCKNRCCGELDFTGYIESPVNPNYPANTEC 835
 Db 359 QPDRQNFCSRQCGNTSTDPDGSITVAQCKNRCCGELGFTGYIESPVNPNYPAGVEC 418
 Qy 836 TTTINPPEKRLIIVPELFLIEDDCGYLVNKRSSNSVTYTTCTYRPAFTSR 895
 Db 419 IWNINPEPKRLIIVPELFLPSEDECGVLNKRSSNSITTYETCTYRPAFTSR 478
 Qy 896 SKKLMQFQKSNQNSARGQVPTVDEDEYQELIETIVRDLVASENHQELIKKKLTK 955
 Db 479 SRKLMNFKTSNARSRGQIPIVYTYDEYEQVELIYVADGRLVASENHQELIKKKLTK 538
 Qy 956 ALFDVLAHPQNYFKYTAQSSREMFPRSFIRLSKVSREFLRPEYK 999
 Db 539 AFEVLAHPQNYFKYT-EKHKEMLPKSFILKLSKVSFLRPEYK 581
 RESULT 9
 Q9TVQ2 PRELIMINARY; PRT: 1664 AA.
 AC Q9TVQ2;
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
 DE Y64G.OA.7 protein.
 GN Y64G.OA.7.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae.
 CC Rhabditidae; Pezodactylidae; Caenorhabditis.
 NX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.

RA Mortimore B.J.;
 RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
 RN 121
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99069613; PubMed=9951916;
 RA none;
 RT "Genome sequence of the nematode C. elegans: A platform for
 RT investigating biology";
 RL Science 282:2022-2058(1998).
 RN 31
 RP SEQUENCE FROM N.A.
 RA Ainscough R.;
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AL117206; CAB60454.1; JOINED.
 DR EMBL: AL110498; CAB60454.1; JOINED.
 DR EMBL: AL110498; CAB57911.1; JOINED.
 DR HSSP: PC0736; IAPQ.
 DR WormPep: Y64G10A.7; CE24549.
 CR InterPro: IPR000152; Asx_hydroxyl.
 CR InterPro: IPR001891; EGF_CA.
 CR InterPro: IPR006239; EGF-like.
 CR Pfam: PF00008; EGF; 25.
 CR PRINTS: PR00311; ECHLAIN-N.
 CR SMART: SM00179; EGF_CA; 4.
 CR PROSITE: PS00310; Asx_hydroxyl; 4.
 CR PROSITE: PS00022; EGF_1; 22.
 CR PROSITE: PS01186; EGF_2; 24.
 CR PROSITE: PS01187; EGF_CA; 3.
 KM EGF-like domain.
 SQ SEQUENCE 1664 AA; 179279 MW; A69F093B4C705832 CRC64;
 Query Match 15.6%; Score 875; DB 5; Length 1664;
 Best Local Similarity 25.5%; Pred. No. 1.9e-68;
 Matches 276; Conservative 99; Mismatches 376; Indels 330; Gaps 45;
 55 DCEADALCON-----TPTSYK-----CCKFGQYQEGROCE-DIECGNEJNGCVHDCN 104
 Db 78 NCSAD-LCHNOGTCVPSFENDNEQVCCPVGF--TGAKQDQDAECCMAK-NGGEHSCVK 133
 Qy 105 IPGVYRTCTCFDFMLADGHNCLDVEDCELENGSCCHTCVVMQSYECCKEFGFLSDNQ 164
 Db 134 TIGVYRCMGHFFLSDCNCTCSIDCAVNSNGCSDRCVNSPQGFQDCDPSDLYLADG 193
 Qy 165 HTG-----HSEEGLSGCKKDCSHI CKEAPRGS-VACSCRPGEFLAKKQDC-- 213
 Db 194 RTGSGGFHFENLI.LKVTSCSTDNGGCEHECENDSNGEYRCRCPVGFKLSENKSCOP 253
 Qy 214 ILTNHNGCCGCHSCDDTAQPECSCHPQYKMHITDQSCLERETVLEVESNTTSVVDG 273
 Db 254 VDPFELKGGQCHCTNNHGRAGQCCYGFELFLSTDRSSCVYIDCAKNGCEHFCENAVKG 313
 Qy 274 DKRYKR-----LMETCAVNGGCDRTCKDTSTGVH-CSPVGFLLQDQK 319
 Db 314 TYRCKCEGVQLSGDRTCEMLGGCCQVNGGCGHCDYDQPDGSHVCKRNGYLLADQK 373
 Qy 323 TCKDIDRCQTPNGSCGFCKNIVGSPQCGKKGFKLTDKSCQDVDCSTDR-TCCHSC 378
 Db 374 LCHDINECHENGGCSQICVNLASGVCEQCQPFRLWKDRKTCEDISEGSNNNGCCEQIC 433
 Qy 379 INHGTFAACNRYGLTYGFTH-CGDINFGSINNNGCCQCVN-VGSEYECGCHRYGLHW 437
 Db 434 SNQGGYWCSEPFELSEDSGSHCHMNBCLINNGCAOLCKRKKGRRCQCFAGYLLAH 493
 Qy 438 NKXDCV-----EVKGS-----LPTSVSPR----- 456
 Db 494 DEKSCVAASDADIFSNIDIEDYSKVPGLDSIDEVTSIESVPADSESRPLVFGRRHYKA 553
 Qy 457 -----VSLHCGSGSGDGGFLRCHSGHLSDDVTTIRSVTFKLNKGCSTKNALFL- 508
 Db 554 CVNFQGLTSELTFSEFVRTDPSEKCPKGFEGS-----TCQLSCSDQNGKCKSRGSLTS 609

[illegible]

Query Match	Best Local Similarity	15.0%	Score 841.5	DB 11	Length 1574
Query	Subject	Similarity	Score	No. 1.7e+65	Matches 245
Query	Subject	Similarity	Score	No. 1.7e+65	Matches 245
34	POGRMAAPC--EDVDECAQDLDCAHDAACNTPTSYKSCSKPGY--OGSGROCEDID	88			
109	POMSGKPEQEGCLSDVDECAHANGCC--EGPCNTVAGFYGRCPRGYOLGSGGKTCQVDV	166			
89	ECGMBLNGGCHDCLINPGNTRCTCFDGFKAHDSNKLVDCELENNGGONHTCWNVX	147			
167	EC-RAENGCGCHRCVNTPGSYLCSCKPFRJHTGRCLTAISCTLNGGCGHOCVOLTV	225			
148	GSYECCKEFGFLSDNHTCIHRESEGSCYMKHCCHCIKEAPRGSVACRPGFSLA	207			
226	TOHKOCSPRGYOLGSGRQVRRS---PCAEENGSGNHICDEL-RGLAHSGCHGYCLA	280			
228	KKQRC--ILTQNHGSGCGHSCDDTDAGFSCSPROYKMTDGSCLREEDTVLEWBS	265			
281	ADRKTCEDVDCALGALQCAHCCANTGSSFKFCVCHAYELGADGQCYRIS--MEIVNS	337			
286	NTSVVDGDKVKRRLMETCAVNNGGDRCTCKSTGVHSCSPRGFTQLDGGKCKSDID	325			
338	-----CEANGGCSHGCSHTSPRLCTCPRGYELDEQKTCIDID	377			
326	ECOTRMGGCHFCKNIVGSPGCGCKKGFRLTDEKSCQDVDECSLDR-TCOHCSTNPGT	384			
378	DC-ANSPCCQACANTPGGYBOSCAGRRLATDGGCGEDVDCECAHGGCHGCSNLAGS	436			
385	FACACNRYTL-----YGFTH	400			
437	FOCFEAGYRDEDEDRGCTSLBESVYDLDRLPFYRPPHVLVRLDELPLRLQDDYGAEE	496			
431	-----CGD---TNECSI---NMG---CQ	415			
407	EAALAEELNGEITTEKPYCLDHSFCHDCSLTDDCRNGSTCPGDDCCCPGWTGIIION	556			
416	QVC-VNTVG---SYECQCHPGYKHMNKKDCVEYKGL--LFTSVSPVSLR---CGXSG	465			
557	ECPCPDYTGKKNSSPCTQ---NCGTCDPLVAGRCRCPGVS---CAHGEDCGPKGF	606			
466	CGDGGFLRGCSJHLSSCVTTIRTSVTKLNEGKC-SLKNAALPFBGLRALPAPFKHSSVK	524			
607	YCKHCRKKCKCHCA-----NRRGHRLYGALCDPGLYG-----	638			
525	ESFRVWNTGSSGKQVPAQPRSPTRPKXETIVBELFETNCKEVTASCTLGCYKRETKR	584			
639	--RICHACFPMAWPGC-----GSEDLIC-----	663			
585	LKLAIRLRLKVAHREDOFHLQSGXNLDVAKKPRFTSERQABSCGVGGGAENQC-VSCRA	643			
661	-----EQSH-----TSCPNKDGSCSKAGFQSGRCAECES	692			
644	CTYTGABERCTL-----CRNGTF-ONEGGMTG--	671			
693	GFFGGCHRHFRCTQCPVACDPVSGECRTQCPRGYQEDCGCECPYGTFFVNSGSCSCVIG	752			
672	EPCCP-----PENSALTKPEAMNNSGCGGLCORGEYSADGAFQCLDGLG-TFOPE	722			
753	AECHVITDECLCPKTKG-----EDCGADCPREGRMIGSGEIDICACSHGASCNBE	802			
723	AGRTSCPPCGGLATKHQCATSF---ODCETRVOCSPGHVYNTTHRCT-----	768			
803	TGTCTCLP--GVSAGRCDTCSAGWYGCGCQIRACACANDGHCDDPTTGRSCAPGWTGLSC	860			
769	--RCVGTGYQPE-----FGKNKCVSGPFG	789			
861	CRACSGHMGPCIHPCNCSAGHNGCAVAG	891			


```

RESULT 11
ID Q9V589 PRELIMINARY: PRT: 1409 AA.
AC Q9V589
DT 01-MAY-2000 (EMBLrel. 13, Created)
DT 01-OCT-2002 (EMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (EMBLrel. 23, Last annotation update)
DE CG7526 protein.
GN CG7526
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophiliidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amaralides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.P.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Suton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blake R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abrita J.F., Agbayani A., An H.-U., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Brannetti D., Bolishevsky S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durkin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Gidek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hoschi D., Houston K.A., Howland T.J., Wei M.-H., Ibeagwa C.,
RA Isalali M., Kalush F., Karpen G.H., Ke Z., Kennison C.A., Kechum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Mervulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Paltan K.A., Pan S., Pollard J., Puri V., Reese W.G.,
RA Reiterer K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Spindling A.C., Stapleton M., Skupski M.P., Smith T.,
RA Splet E., Spindling A.C., Stapleton M., Strong R., Sun E.,
RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri U.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195 (2000).

```

```

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Miya S., Crosby X.A., Matthews S.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,
RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celinker S.E.,
RA Clamp M., Dysdale R., Emmert D., Fries E., de Grey A., Harris N.,
RA Krommler B., Marshall B., Milburn G., Richter J., Russo S.,
RA Seale S.M.J., Smith E., Shu S., Soutnik F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT "Annotation of Drosophila melanogaster genome.";
RN [4]
RP SEQUENCE FROM N.A.
RA Adams M.D., Celinker S.E., Gibbs R.A., Rubin G.M., Venter J.C.;
RN [5]
RP Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RA F.Y.Base;
RN [6]
RP SEQUENCE FROM N.A.
RA Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
ER EMBL: A203558; AAF50538.2; -
CR HSP: P00736; APQ.
CR FLYBase: FBgn0035798; CG7526.
DR InterPro: IPR000152; Asx hydroxyl.
DR InterPro: IPR001887; EGF_Ca.
DR InterPro: IPR006239; EGF_like.
DR InterPro: IPR003410; Hyalin.
DR InterPro: IPR006210; IEGF.
DR InterPro: IPR000436; Sush1_SCR_CCP.
DR Pfam: PF00008; EGF_11.
DR Pfam: PF00084; Sush1.
DR SMART: SW00332; CCP.
DR SMART: SW03181; EGF_13.
DR SMART: SW03179; EGF_CA.
DR PROSITE: PS00010; ASX_HYDROXYL.
DR PROSITE: PS01186; EGF_2; 3.
DR PROSITE: PS01187; EGF_CA.
DR EGF-like domain.
SC SEQUENCE 1409 AA; 15413 MW; F9ACDE179AF354C CRC64;

Query Match 75.5%; Score 775.5; DE 5; Length 1409;
Best Local Similarity 23.6%; Pred. No. 1,1e-59;
Matches 212; Conservative 91; Mismatches 296; Indels 230; Gaps 26;

43 QEYDCAAGDDBCHADALCQNPYSKSGKPGYGBEGRCEIDECNEJNGCWHQC 102
376 EKVEKCSG-----TCLASEDNISFSCICPTGYRSEAFSCQIDDECAEDTH-LCSHTC 428
103 LNPGRYR-----CTCFPGFMLADG 123
429 QNPRGIVGCGCGELNLVEFYTCLENLCEVNNNGCEQICLTARSGVCACREGFRLSADG 488
124 ENCLVDECLENNAGGQHTCVVMGSEYECCKEGF---FLSDNQHITIRSEBGLSKMK 180
489 KSCEDVDECLVNNGGCGQVCRN.PSGYSGICAAGYELLKLDGIPGCFIDE-----CSGR 544
18: DRGCSH--LCKAFRSVACGCRPGFEFLAKNRDCLT----- 216
545 THGCSQMLC-ENLNSSTYTCIDPGYALGLDNH--IVSLNSSTFTDSTSSSTPSAHTCL 601
217 ---CHNGGCGHSCDPAADPEGCHGQYKMHHTDGRCLREDTFLBVTESNTTSVVD 272
602 DIDEGLANGNSHFPONRPGFGACAPGLVLSERKTCQ----- 643
273 GGRKVRRLMETCAVNNNGCPRCTKDTSTYHSCSPVGFILQLDQKTKDIDECQTRNG 332
644 ---DECLDSNQCQSLCINPGGPRACACETGFEFLPFDGFCADIDECQDVG 693
333 GCDHFKNTVGSFDCGCKKGFLLDDEKSCQOVDEG---JDRTCCHSCINHGTFRCA-- 388
694 NSDIDCINLSTHACACERGYELANDKLSCLVDDEAGLSGGCSHCKINKAGTFEGCP 753
389 ---NRGTYLVGTHGCTNCSINNGC 414

```


OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99350211; PubMed=10419698;
 RA Yang Q., Oca K., Tian Y., Kumar A., Wada J., Kashihara N., Wallner E.,
 Kanwar Y.S.;
 RT "Cloning of rat fibrillin-2 cDNA and its role in branching
 morphogenesis of embryonic lung."
 RL Dev. Biol. 212:229-242(1999).
 DR EMBL: AF150560; AAC34439.1; ..
 DR HSP: J35555; 1EXN.
 DR InterPro: IPR002086; Aldhyde_dehydr.
 DR InterPro: IPR000152; Asx_hydroxyl.
 DR InterPro: IPR001881; EGF_Ca.
 DR InterPro: IPR001438; EGF_11.
 DR InterPro: IPR006209; EGF_1like.
 DR InterPro: IPR002212; Fibril-assoc.
 DR Pfam: PF00683; TB; 9.
 DR Pfam: PF00683; TB; 9.
 DR PRINTS: SM0010; EGFLOOD.
 DR SMART: SM00179; EGF_CA; 42.
 DR PROSITE: PS00073; ALDEHYDE_DEHYDR_CYS; 1.
 DR PROSITE: PS00010; ASX_HYDROXYL; 43.
 DR PROSITE: PS00022; EGF_1; 2.
 DR PROSITE: PS01186; EGF_2; 36.
 DR PROSITE: PS01187; EGF_CA; 43.
 KW EGF-like domain.
 SQ SEQUENCE 2906 AA; 313371 MW; 9EB64E727044EF58 CRC64;

Query Match 12.8%; Score 720; DB 1; Length 2906;
 Best Local Similarity 26.9%; Pred. No. 2,7e-54;
 Matches 256; Conservative 96; Mismatches 308; Indels 292; Gaps 54;

QY 45 DVDECAAGJDDCHADALCONTPSYKSCSKPGYQ--ESQCEDIDECGNELNGSCVFCDC 102
 Bb 1192 DINECSLS-DNLCRNGKCYMMIGYQSCSNPGYQTPROGCSDDIDEC-MKNGSCDTQC 1249
 QY 103 LNIPTNYRCTCFDGSFLAHGHNCLDVECLENN-----GGQHTCVVWWSYECCKEG 157
 Bb 1250 TISEGSEYECSCSEGAALMPDGRSCADIDEC-ENNDIDDCG---QCTNIPGEYRLCTDGG 1305
 QY 158 PFLSDNHTCHIRSEEGSCMKKHGCHICK---EAPRGSVACERPGFELAKNGRDC 213
 Bb 1306 FMASVDMKTCIDVNECDLN-----FNICMFGECECTKGSFICHQQLGYSVKGATGC 1357
 QY 214 --ILTCNHNGGCG--QHSCDDTADGPESCHPQYKMHNDGRSCLERETVLEVESNTTS 269
 Bb 1358 TDVDECEIGAHNCDMHASCLNVPGSFKSCREGKV--GNGIKCIDLDJSC-----ANGT- 1408
 QY 270 VVDGDKRYKRLMETCAVNNGGCDRTCKDTSTGVHSCPGVFTLQDQKTCXDIDECQT 329
 Bb 1409 -----HCGSIN-----AQCNTPGSYRCACSEFT--GGFTCSVDECAE 1447
 QY 330 RNGGCDH-FCKNIVSPFCGCKKFKLLTDEKSCQDVDECSLERTCDH-SCINHPGTAC 387
 Bb 1448 NINLCENGGCLNVPAAYCECEMGCTPASDSRSQDIDECGFCNICVGTGNNIPGMFHC 1567
 QY 388 ACRNGYTL-YGFTHGDNMGES--IN--NGGCGVCVNTVSGYECQCHPGYKLLHNNKDC 442
 Bb 1508 ICDDGGLDRTGCHCTDIDECADPNCVNG---LCVNTPGRECKCPDQQLVATGVGC 1563
 QY 443 VEVR-GLLPTSVSPVSLHCGKSGGDDG---CFLRHGHGIR-SSDVTFTIRSVFKLNEG 499
 Bb 1564 VDNRYGNCYLKRGPR-----GDSLSCKTEWGVGVSSSSCC----- 1599
 QY 499 KSLKNAELFPGGLPALPBRKSSVKSFRYNNLTGSSGKVPAGAPSP-PRKEXFTVE 558
 Bb 1600 -CSJGKAMGNPCETCPPV-----NSTEYYSL-----CPGEGG--FPPNQITLIE 1641
 QY 559 -----FELETNQ-----KEVTASCO----- 573
 Bb 1642 DIDECCELPGLCGCGNGCINTGSPQCECPGYLSEET-RIGSDIDECFAHRAVCGPGTGY 1701

QY 574 -----TSCIVKTRERLRKAIPTLRKAVRBOFHQLQSGMN-LDVAKK----- 615
 Bb 1702 NTLANTTICLPP-----YMGVNGHNMOMRKRCFCVSYNGTTC 1741
 QY 616 ---PRTERQAESGVGQGAENQ-CVSC-RAGTY----- 646
 Bb 1742 ENELPNVTVKRC-CCYVWYGRKAMRPFCECPPTGADFTICGNIPIGFTFDIHGTCAVD 1800
 QY 647 YDARERCLCPVGTQNEGGQTCPPCRPGNSGLKTPKAN-----XSEGG--- 695
 Bb 1801 IDCKEIPGICANGVCINQGSFRC-CP-----TGFSYNDLLVCEDIDECNSGD 1850
 QY 696 GLCO-----PGEVSADGFAPQCALG-TFQPAAG--RTSCFP-----COGLATW 739
 Bb 1851 NLTCRNADCCNPSGRFCE-----CAAGFKLSPYACVDRHCLSEINVCSSGLCVLD 1903
 QY 740 CGATSPQ-DETRVQCSPGHFYNTTHRCIR--CPVGYQPEFGKXKCVSCPG-----N 790
 Bb 1904 CG--SYQCIKNGSKASQCTMCKEYDECEHRPCNGTCKNTVGSYNCLCYPGFELTHNN 1961
 QY 791 TTDFPGSTNI--TCKNRCGSELCP-----TCYTESPNYGNVPAATEC 835
 Bb 1962 DCLDIDECSSFFSGVCRNGRCENEISGFKCLNBSYEELPPGXNCTDTN-EC 2012

RESULT 14

Q96CJPa
 ID Q96CJPa PRELIMINARY; PRT; 2809 AA.

AC Q96CJPa
 DT 01-DEC-2001 (TREMBLrel. 19, Created;
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Fibrillin3.
 GN KIAA1776.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteiostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 CX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=21245130; PubMed=1147906;
 RA Nagase T., Nakayama M., Nakajima D., Kikuno R., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. XX.
 RT The complete sequences of 103 new cDNA clones from brain which code
 RT for large proteins in vitro.";
 RL DNA Res. 8:85-95(2001).
 DR EMBL: AB053450; BAB47408.1; ..
 DR InterPro: IPR00152; Asx_hydroxyl.
 DR InterPro: IPR001881; EGF_Ca.
 DR InterPro: IPR006209; EGF_1like.
 DR InterPro: IPR002212; Fibril-assoc.
 DR Pfam: PF00008; EGF; 45.
 DR Pfam: PF00683; TB; 9.
 DR SMART: SM00179; EGF_CA; 41.
 DR PROSITE: PS00010; ASX_HYDROXYL; 41.
 DR PROSITE: PS00022; EGF_1; 2.
 DR PROSITE: PS01186; EGF_2; 36.
 DR PROSITE: PS01187; EGF_CA; 40.
 KW EGF-like domain.
 SQ SEQUENCE 2809 AA; 300323 MW; 20C04CC006C0161F CRC64;

Query Match 12.7%; Score 710.5; DB 4; Length 2809;
 Best Local Similarity 24.8%; Pred. No. 1.8e-53;
 Matches 262; Conservative 79; Mismatches 316; Indels 401; Gaps 50;

QY 45 DVDECAAGJDDCHADALCONTPSYKSCSKPGYQ--GEGRCQEDIDECG----- 91
 Bb 1071 DVDECARPPLCGG-CTGNTDGSYKCGCPGHEZAKTKACEDIDECSLSGLCRPHGCG 1129
 QY 92 -----NET-----NGCCHDCNLIPGNVRCCTPDGFWLAH 121

```

Db      1130 VNVIGAFCCSHAGFQSTPRQGVCDINECVQNGCCDHCINTESSYRCSGCGVSLMF 1189
Cy      122 DGHNCLDVDECLENNNGG--CHTCVNVWNGSYECCKEGFPLSDNCHTCIHSSEGLSCMK 180
Db      1190 DGRACADVDCEENPRVCDQGHCTNMGGHRCLODYDFMATTDMRRTCVDDVE-----CDLN 1245
Cy      181 DHGSHITKEAARPGSVACEPRPFELAKNORC--ITCHNGGCGCH--SCDCTADGPE 236
Db      1246 PHICLHGDCEHTKSPVCHCQLGYVVKGATGCDVDECEGCHNCDSHASCINTIPGFS 1305
Cy      237 CSHPOYKMTDGSCLEREDTVLEVSNTTSVVDCKXKXKRLMLMETCAVNNGGDR 296
Db      1306 CRCLPQWV--GDGFEBCHDLDEICISQEHRCSPR-----GD----- 1337
Cy      297 CKDTSTGVNCSGPFYTLQDLGKTCCKIDECQTRNGGCDH--FCCKIVGSFDCGCKKFKL 355
Db      1338 CLNVPGSYRCTCRQGA--GDGFCEBDECAENVDDICDNGCGANAGVRCEGEMGFDR 1395
Cy      356 LTDEKSCDVDECSLDRTCDH--SCINHPGTFACACNGYTL--YGFTHCGDTNCS--IN- 430
Db      1396 TEDHACQDVDECAQGNLCARFSCENLPGRFCINQGYELDRGGGCTDINECADPVNC 1455
Cy      411 -NGGCGQVCVNTVSYECQCHPGYKLMNKKDCVEVAGLPTSVSPVSLHCGSGGDDG 469
Db      1456 ING-----VCINTPSSYLCSGQDFELNPSGVGVDTR-----AGN 1491
Cy      470 CFLRCH-----SGIHLSD--VTTRTSYTFKLNKGKSLKNAELFPGLPPLPEKHSV 523
Db      1492 CFLETHPRGDSGISCSAIGVGTBRASC-----CSLGRAMGNPELCP-----M 1536
Cy      524 KESRRYVNLTCSSGQVGVAGP--RPSTPKEMPTVEPELENNQEVATASCD--LSCITKRT 581
Db      1537 ANTEYERL-----CPGGEGFQPNR-----ITVLEIDIECQELPGLCQSGSPCV--- 1580
Cy      582 EKLRKAIRLTKAVHBPQHLQLSGNVLVAKKP--RTSE-----RQASCGVGGHA 634
Db      1581 -----NTBSPFCCEPFGYHLSEHRIREDIDECSTHSG-- 1614
Cy      635 ENQVSCRAGTYVDGARERCILCP-----NCTFQNE----- 665
Db      1615 -----ICPGTCVNTLGNVTCFPAEY--QVANGNNCMOMKSKVCFRRYNGTCQHELAFNV 1669
Cy      666 -----EGQ--MTCEPQPRP-----GNSGA--LKTPEAMNYSGG-- 695
Db      1670 TRKMCCSGYINIGLAMPCEACPTIPSPDYQILCGNQAPGFJTHFGKPLDIECGSEIP 1729
Cy      696 -----GLC--QPEYSAD--GF-----APCQ----- 712
Db      1730 AICANGICINQISFRCECPAGFVYNSILLACEDVDEGCSHESPCCQANADCNTPGSYRC 1783
Cy      713 -----JCALGTPQPEAGRTSCF-- 723
Db      1790 KOTRSYKLSPGACAGVQNECREIPNVCSHGQCKPTESYMCJCHRG--FQASADTLCMDI 1843
Cy      720 -----PCGGG-----LATRQAGATSPQCCETRV--QCSGPHFYVTT- 763
Db      1849 DECDRQPCGNGTCKNIISVNCLECPGQVVTNMGCCVDFDCTLTVGQVCRPGCLNTAG 1908
Cy      764 -----THRCI--RCPVGYOPEFGKNNCVSCPNTTTDFGS 798
Db      1909 SFHGLCCQGFELTADKKACVDTNECLSLAGTCLPOTCONLBSFRCI--CPGPGVQGFHC 1967
Cy      799 TMITQCKNR-----RGGELGDF-----TGYESRN 824
Db      1968 IDIDECSEEPNLCFGTCTNSPGSFQCLCPGFSVLSDR 2005

```

RESULT 15

CNC435

ID 08C435

PRELIMINARY:

PRT:

810 AA.

AC 08C435:

DT 01-MAR-2003 (TREMBlrel: 23, Created)

01-MAR-2003 (TREMBlrel: 23, Last sequence update)

```

DT      01-MAR-2003 (TREMBlrel: 23, Last annotation update)
DE      Matilin 2.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
CX      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxId=10090;
RN      (1)
RP      SEQUENCE FROM N.A.
RC      STRAIN=C57BL/6J; TISSUE=Hippocampus;
RX      MEDLINE=22154683; PubMed=12466851;
RA      The FANTOM Consortium,
RA      the RIKEN Genome Exploration Research Group Phase I & II Team;
RT      "Analysis of the mouse transcriptome based on functional annotation of
RU      Nature 420:563-573 (2002).
CR      EMBL: AK083151; BAC38785.1;
SQ      SEQUENCE 810 AA; 90072 MW; 12CB0D391B6C27BC CNG64;

```

```

Query Match      12.4%; Score 695; DB 1.; Length 810;
Best local Similarity 30.0%; Pred. No. 8e-53;
Matches 160; Conservative 71; Mismatches 224; Indels 78; Gaps 14;

Cy      62 CCKPTSYKSGCKPGY--GSGRCCED:DECGNELNGGCVHDCNIPGVNCTCFDGM 119
Db      253 CUNTSGSYICKKQGGYISTQKTCRIIDLQATE--DHCCQGLCVNMLGSPVCCQVSGYTL 312
Cy      120 AHGHNCLDVDECLENNNGGCHTCVNVWNGSYECCKEGFPLSDNCHTCIHSSEGLSCMK 179
Db      312 AEDGRCCAVDYCASERNGCHECVNASSYLCHREBFALNSCKTKSLD---YCAS 367
Cy      180 DHGSHITKEAARPGSVACEPRPFELAKNORC--ITCHNGGCGCHSCDCTADGPE 237
Db      368 SNHGQHECVNA--QTSACRCLKGMLEPDKTORINYYCAANKPGCHECVNTEGHYC 426
Cy      238 CSHPOYKMTDGSCLEREDTVLEVSNTTSVVDCKXKXKRLMLMETCAVNNGGDR 296
Db      427 KCRQGVNLDPRKCTSRVDHCAQCHGEOLCLNTEESFVCCSSGFLINDLTKCSAD 486
Cy      285 TCAYNNGGDRCTCKDSTGVNCSGPFYTLQDLGKTCCKIDECQTRNGGCDHFCCKIVGS 344
Db      487 YLLSNHGGEYSVANTDKSPACQCEGHVLRADGKTCAXKLDSCALGDHCEHSVSSDS 546
Cy      345 FCGCKKGRKLTDEKSCQVDEEC--SLRTCDHSCNHPGTFACACNGYTL--YGFTHCG 402
Db      547 FVCCQPEGYILRCDGKTCRRKVCQCVVNHGSHLCVNSGESYVCKLEGFRLAEDGRCK 606
Cy      403 DNECSINNGSCQVCVNTVSYECQCHPGYKLMNKKDCVE----- 444
Db      607 RKNVCKSTQHGCEHNCVNNNSYLCRCEGFVLAEDGHCRCRTEGPIDLVFVIDGSKS 666
Cy      445 -----VKGLFT--SVSPVSLHCGSGGDDGFLRCHSGIHLSSDYTTTRTSV 492
Db      667 SEENETVNHFTVGTGLDLSLAVSPKA-----ARVGLQVSTQVRIET 708
Cy      493 FK-----LMSGKSLKNAELFPGLPPLPEKHSVKSFRVYVNLTCSSGKQVP 541
Db      709 JGFSASAKEMKAAVTHKKYKGSMTGLALGH--MERSFTQVEGARPSQVVP 760

```

Search completed: October 22, 2003, 11:30:17
 Job time : 130 secs